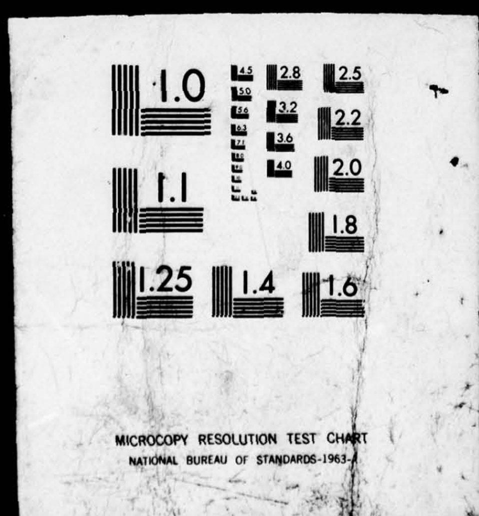
AD-A036 030 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 5/2
ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION --ETC(U)
DEC 76

UNCLASSIFIED AGARD-CP-207 NL

1 OF 2
AD
A036030



VII. CONCLUSIONS.

The minicomputer is a viable and desirable alternative to a mainframe computer in several

AGARD-CP-207

AGARD-CP-207

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No. 207

on

Advancements in Retrieval Technology as Related to Information Systems

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY ON BACK COVER

AGARD-CP-207

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

9 Conference proceedings.

6 AGARD Conference Proceedings No.207

ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION SYSTEMS,

11 Dec 76

12 156p.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Arlington, VA, USA, 20–21 October 1976.

400043

THE MISSION OF AGARD

The mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

– Exchanging of scientific and technical information;

– Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

– Improving the co-operation among member nations in aerospace research and development;

– Providing scientific and technical advice and assistance to the North Atlantic Military Committee in the field of aerospace research and development;

– Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field;

– Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

– Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Program and the Aerospace Applications Studies Program. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced directly from copy supplied by AGARD or the authors.

Published December 1976

ISBN 92-835-0183-7

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London, W1P 1HD

THEME

The ultimate test of any information system is whether or not the user can quickly, easily and comprehensively obtain relevant information stored therein. As data bases expand and multiply, retrieval technology must advance if the user is to benefit fully from this increased information potential. For this reason, three of the four sessions of this meeting, covering minicomputers, bibliographic data base sharing and future technological advances, explore in some depth areas critically important to the user's need to access an increasing number of scientific and technical information collections with greater ease and efficiency.

The other session was somewhat differently oriented and addresses numerical data bases, a relatively new and rapidly expanding addition to information systems. Presentations include not only the development of such data collections but several applications as well.

Δ Partial

CONTENTS

THE ROLE OF THE MINICOMPUTER IN THE INFORMATION RETRIEVAL BUSINESS

B.K. Penney

Department of Computing and Control,
Imperial College of Science and Technology,
London SW7 1LU
England.

SUMMARY

Various aspects of the use of computers in information retrieval are examined to determine which areas may be better or more profitably served by using minicomputers rather than the more usual mainframe system.

I. INTRODUCTION

Before examining the different aspects of information retrieval there is a discussion of the difference in characteristics between minicomputers and maxicomputers (also called mainframes). The first aspect of information retrieval to be discussed is the computer network; in particular, the communications subsystem of it. Next is the data entry process; this is fundamental to the creation and maintenance of a data base and minicomputers have many features which make them suitable for controlling this task. The information retrieval system itself is discussed, both on-line interactive versions and off-line or batch types. One specific application considered is the use of a minicomputer to replace a simple terminal connected to one or more remote data bases via a network. This approach, although involving higher initial expenditure, does confer some advantages. There is also some discussion of the use of distributed systems.

II. CHARACTERISTICS OF MINICOMPUTERS AND MAXICOMPUTERS.

1. Introduction:

Minicomputers and maxicomputers have characteristics which differ from each other strongly in some respects. Most attempts to define a minicomputer start with discussions of maximum memory size, disk storage, software and other factors, and finish by making a definition based solely on cost. As it is not the intention in this paper to become involved in these usually fruitless discussions, I will not attempt a definition but will review some of the differences between minicomputers and maxicomputers.

2. Cost:

The most striking difference is cost. A minicomputer with disk, line printer, teletype and 64K bytes of main store will cost in the region of £20,000 - £30,000. Amortized over 5 years and taking into account service charges, supplies, etc., this would represent a monthly cost of £500-£800. A mainframe system, say a middle size IBM 370, would cost from £0.5M upwards. Amortizing this cost over five years, and adding all associated costs of staff, maintenance, building, air conditioning, supplies, etc., the corresponding monthly costs would be upwards of £20,000. Of course, the big machine has many more resources than the smaller machine and supports more users. The basic question is whether one has a better deal using all the resources of a small machine or a small share of a big machine.

One of the more striking facets of cost which support use of big computers is the cost of disk storage. The cost per bit of disk storage decreases as the size, or capacity, of a disk pack increases. The typical maxicomputer disk system will be several times cheaper in terms of cost per bit than the typical minicomputer system. It is not simply a case of transferring the big disk systems onto minicomputers. The computer manufacturers do not, for marketing as well as for technical reasons, support the use of the bigger disk systems on the smaller computers.

3. Software:

One of the major differences between minicomputers and maxicomputers is the provision of software. Minicomputers have, for example, traditionally been supplied with comparatively primitive operating systems, often just a single-user and, possibly, a foreground/background system. In very recent years this situation has been improving and several minicomputers do now have simple-minded timesharing systems. Programming languages for minicomputers are mainly restricted to Assembler, Fortran and Basic or one of its variants, and often to subsets of these only. These are not necessarily the best languages available for developing information retrieval systems. The third area where software is lacking for minicomputers is in data base management systems where often only a primitive filing system is provided. The reason for these deficiencies is mainly financial. The table below shows the annual revenue for some of the major computer manufacturers. It may be seen that the largest of the minicomputer manufacturers, Digital Equipment Corporation, has a revenue some 30 times less than the largest of the mainframe manufacturers, IBM. As a majority of the cost in designing and implementing computer systems lies in the software, it is clear that there are nowhere near the same financial resources available to develop minicomputer software as there are for maxicomputers.

ANNUAL SALES

	$ MILLION	
IBM	14,400	
Univac	1,430	
Control Data Corporation	1,363	
Honeywell	1,324	
NCR	950	
Digital Equipment Corporation	480	Mini
ICL	427	
Hewlett Packard	165	Mini
Data General	120	Mini
General Automation	75	Mini

Sources: Datamation September 1976
Mackintosh Consultants, Microcomputer Minicomputer Survey, London, 1976.

4. Program development:

Program development on minicomputers also differs in a number of respects. Overall, the limited hardware and software resources limit the speed of development, particularly the edit-compile-load-execute cycle. Also, the more limited subroutine and program libraries may require more software to be written on the minicomputer than on a maxicomputer with more extensive libraries.

On all but the larger minicomputers there may be insufficient resources to allow testing and development to proceed concurrently with production running.

5. Availability:

The availability of any computer system is limited by the amount of time scheduled for systems maintenance and development facilities. Most big mainframes require several hours of dedicated prime shift time per week for these activities, whereas most minicomputers require practically no time at all after initial development.

6. Reliability:

All computers are likely to have faults in both hardware and software. Hardware faults are usually self evident inasmuch as the computer stops, or, for example, a disk suffers a head crash. Software errors constitute a bigger problem as their effects are not always immediately self evident and are thus harder to locate and correct. As the size and complexity of software increases, so do the difficulties of ensuring its correctness, of maintenance and of development. It often happens that big operating systems are released containing as many as a thousand errors. Simply by virtue of its small size and relative simplicity, reliability is potentially better on minicomputers than on big mainframes. Mainframes tend to have better hardware fault detection features (e.g. the IBM 370 series) than on minicomputers, which helps to reduce consequential down time. With software, however, there are no corresponding error detection facilities and the minicomputer is intrinsically in a much stronger position. It is not uncommon for minicomputers to run for months without a system crash, while with mainframes, system faults at a rate in excess of one a day are not uncommon.

7. Personnel:

Systems software programmer support on a minicomputer may be as little as one person (or none at all) as minicomputer vendors tend to have operating systems that are relatively static. Contrast this with the typical big machine installation where the incorporation of software patches, updates, and new releases are a continuing activity throughout the lifetime of the system and require many personnel to perform these tasks. In addition, mainframe systems commonly have one or more resident engineers to maintain the hardware.

III. NETWORK COMMUNICATIONS.

1. There are a large number of national and international computer and terminal communications networks in the world today. It is generally accepted that the communications systems should be driven by minicomputers and that any users should be connected to the network via a standard minicomputer (known generally as a network processor) provided by the network. This philosophy has been followed from the beginning of modern history (ARPANET (1)) to the latest networks such as the new European EIN (3) network. In ARPANET (2), for example, there are two types of minicomputers used as network processors. The IMP (Interface Message Processor) can interface up to four geographically close **HOSTS** (user computers) to the network, and the TIP (Terminal Interface Processor) can interface up to 63 terminals without those terminals needing to be attached to a user computer. These were originally based on Honeywell DDP 516 minicomputers. Figure 1 shows a schematic of an ARPANET type scheme. The reasons for this approach are discussed in the following paragraphs.

2. Minicomputers are of comparable power to mainframes at handling character oriented input output, and are thus able to do the job as well. In fact, the minicomputer will be doing only this one job so its performance should be better and more consistent than a mainframe which is doing several other jobs simultaneously.

3. If the communications handling was done in the mainframe, a complete and very complex program would be required for each type of mainframe in the network. The program development would thus be expensive and time consuming and would virtually eliminate networks of dissimilar computers. Putting the communications handling into the minicomputer means that only one type of program is needed for the lower levels of the communications protocol, including store and forward activities, and more effort may be expended in getting it right. There must still be a program in the users computer in order to communicate with the network machine, but that program will be very much simpler than if all levels of network

communication were in the users computer.

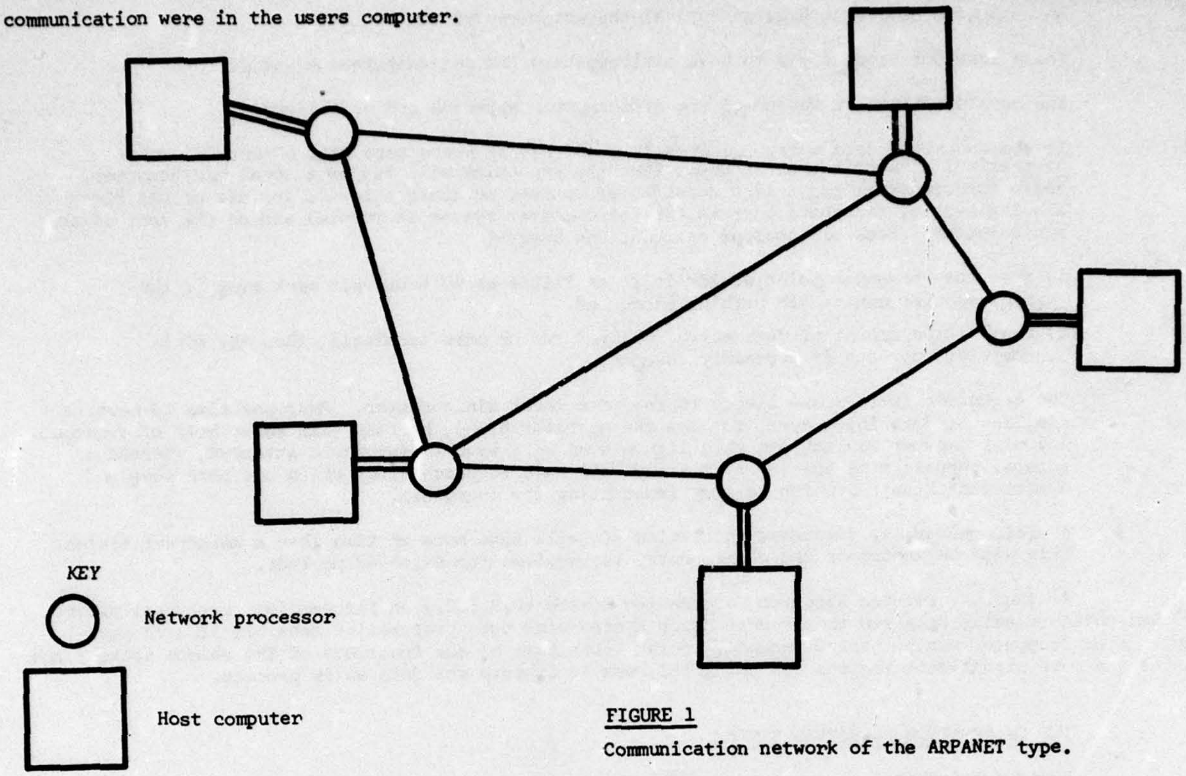


FIGURE 1

Communication network of the ARPANET type.

4. Putting the communications handling into a separate minicomputer reduces the load on the user computer. High speed character communications and message switching is a heavy load on a computer and the required software occupies a substantial amount of main store. Offloading it on to a minicomputer is clearly cost effective as minicomputer hardware is a great deal cheaper than mainframe hardware.

5. The minicomputer isolates the user computer from the communications network, hence any error by, or failure of, the mainframe cannot interfere with the rest of the network. The greater reliability of a minicomputer will, in a store and forward network, give greater reliability for the communications system as a whole as, even if the user computer is not working, the minicomputer (network processor) will still function and pass on messages for other nodes in the network. On ARPANET (1), for example, total up time of an IMP averages 99%, while total up time for a HOST does not generally exceed 90%.

6. Overall this use of minicomputers as network processors seems to be entirely justifiable and sensible, and there is no reason why this situation should change.

7. An interesting exception to this is DEC's new network architecture, called DECNET, (4). In this system different computers are linked together directly without using front end processors. The computers linked, although different (PDP11, PDP10, PDP8), are all made by the same firm and as such this is a restricted case and not a general purpose heterogeneous network.

IV. DATA ENTRY.

1. The creation and maintenance of a data base can be carried out completely off line using paper tape, punched cards or some magnetic medium. The data is keyed on to the medium and transferred to a remote computer to be processed off line as part of a batch stream.

The disadvantages of this approach in terms of handling difficulties (especially paper tape!), error rates and turnaround are well known to everyone who has had to prepare data in this manner. Paper tape is the worst to handle, of course, while punched cards, although easier to handle, suffer from not having the ability to record lower case characters directly without the use of escape codes. This is particularly important for bibliographic data.

2. The alternative to off line data entry is on line data entry. Where an organisation does not have a mainframe in house it can use terminals which are connected either to a remote timesharing system or to a local in house minicomputer.

The arguments for having data entry done on line with an interactive system are that

a) The data can be validated and mistakes corrected straight away.
b) The data can be displayed on a screen suitably formatted and possibly with additional information.
c) No handling of vulnerable media such as paper tape is involved, and hence no errors due to mishandling or damage can occur.
d) Turnaround is shortened. At the end of a session data can be copied to a magnetic tape and reports printed immediately, rather than waiting for a period of up to several days for remote batch working.

e) An automatic device such as an optical character reader can be used.

These considerations apply to both minicomputers and maxicomputers equally.

3. The considerations in favour of the minicomputer approach are as follows:-

a) If the amount of data entry required is such that it needs more than a certain number of terminal connect hours per week, then the economics will favour a local minicomputer. Where the crossover point lies depends, of course, on charges levied for use of the time sharing system, telephone charges (if the computer system is remote) and on the cost of the minicomputer. Back-of-envelope calculations suggest

i) that the crossover point would lie at as little as 10 hours per week even if the minicomputer system did nothing else, and

ii) that if the amount of data entry required two or more terminals, then the minicomputer approach is certainly cheaper.

b) The ergonomic factors are likely to favour a local minicomputer. Response time to service one line of data input must, to keep the operator happy, be less than about half of a second. Any minicomputer can achieve this figure even with several terminals attached, whereas a general purpose time sharing system may well have response times which are both long and erratic; a situation which is very frustrating for operators.

c) A minicomputer, as discussed in section II, will have more up time than a mainframe system. This will be important where data entry is required for extended periods.

At the U.K. Defence Research Information Centre (D.R.I.C.) an information retrieval project (5) has until recently been run on a remote batch system with data preparation done off line on paper tape. A minicomputer was purchased recently to take over some of the functions of the remote system, and one of the more significant reasons for doing this was to improve the data entry process.

V. THE INFORMATION RETRIEVAL SYSTEM.

1. Basic approaches.

An information retrieval system can be implemented using one of three basic approaches.

a) It can use a dedicated mainframe system

b) It can use a share of a general purpose mainframe system, either an in house or a remote bureau system

c) It can use a dedicated minicomputer system.

At one extreme the information system could involve very large databases, extensive usage and, possibly, many interactive terminals. At this extreme the dedicated mainframe system has been used as that has been the only way technically of providing the necessary resources. Such an approach is followed in the System Development Corporation (6) and European Space Agency (7) systems, for example. Each of these systems contains many large data bases and each allows many concurrent users from a large geographical area.

At another extreme the information retrieval system may consist of comparatively small data bases and involve only occasional use. In these circumstances it would be possible to use either of approaches b) or c). The minicomputer costs are fixed and are essentially independent of the amount of usage. In contrast a mainframe system is normally charged for according to the amount of usage involved. For occasional use the fixed minicomputer costs would be higher than the cost of using a mainframe system, and there would thus be economic advantages to using the mainframe system.

In between these extremes are many other possible circumstances, and the approach to be used has to be considered in terms of cost, ease of development, ease of running,performance and reliability. The two approaches, shared use of a mainframe and exclusive use of a minicomputer, are discussed in the following paragraphs.

2. Cost:

If the purchase cost of a minicomputer is amortized over a period of time then the resulting monthly cost is to be compared with rental of time on a bureau machine. McIvor has shown that in one case (8) the comparative costs of running the same application on a bureau machine and on an in-house minicomputer is five to one, a result which so clearly favours use of a minicomputer that one wonders why many more users don't take this approach. It may, perhaps, be because many large bodies, such as universities and defence organisations, provide use of an in house or bureau mainframe machine 'free' to internal users, whereas the internal users would have to put a strong case for acquiring a new minicomputer.

For on-line use, communications charges arise if the terminals are remote from the computer.If the computer used is a minicomputer and the terminals are clustered, then it should be possible to set both in the same place, thereby avoiding telephone costs completely.

An unexpected penalty faced one prospective on-line user (10) in the U.K. who wished to be connected on-line to a private data base for several hours each day. The on-line period was considered sufficiently long by the operators of bureaux that he was asked in effect to buy the disk drive which held his data base. This user did, in fact, decide to implement his system on a dedicated minicomputer.

3. Performance:

For off-line processing the user will not usually be concerned with factors such as speed of processing or operational problems. His main concern will usually be to achieve a reliable, short, turn-around regardless of the time the job spends in the machine. While a minicomputer may well take somewhat longer to run a given job than a maxicomputer, the minicomputer may be able to start processing the job sooner due to the different operational environment. Overall, *for batch type processing, the most significant* factor will be whether the system used is in house or not rather than whether the system is based on a minicomputer or maxicomputer. For jobs which are not run on demand but on a periodic basis, such as monthly SDI lists, there should be no difference to the user at all.

For on-line working the minicomputer might be able to give superior performance to that of a remote mainframe system. This arises from the superior terminal handling properties of the minicomputer when executing only the interactive program. That is, one program with two or three terminals running on a minicomputer will give a better performance than a general purpose time sharing system running a hundred terminals for diverse tasks. If the mainframe computer was dedicated to the information retrieval system then, of course, its greater total resources would inevitably give better performance than would a minicomputer system.

4. Reliability:

For batch running the main effect of the computer being out of action, whether due to scheduled maintenance or to faults, is to increase the delay between the user submitting a job and receiving the output. This might be inconvenient but should not be over disruptive e.g. if a monthly SDI list is a day late that will probably not stop the intended recipient working. In an on-line system the computer being down does prevent the user from working to schedule, and this is likely to be much more disruptive. The greater reliability and overall availability of minicomputer based systems should give a significant advantage over mainframe based systems, especially with interactive working.

5. Security/Privacy:

If questions of privacy (ie unauthorised access) arise, then clearly having the computer system on site and dedicated to that single application is a very great advantage, and obviously a minicomputer is suitable for this. The fact that it requires fewer personnel to develop and operate the system is another advantage.

6. The library information retrieval system is one area which seems well suited to minicomputer use. A library which uses an on-line interactive system needs it to be available throughout the working day. If the level of use is sufficiently high then the cost of using a mainframe system may exceed the fixed costs of using a minicomputer. In a paper on the proposed BLIPS system (9) for special libraries it is shown that a database which accommodates the equivalent of 50,000 library items would require 50M characters of disk space, an amount which can be supplied on one disk pack on modern systems. For example, the CAIRS (10) minicomputer based system is in use by a small industrial research association in the UK and is a current and retrospective information system which offers the usual facilities such as SDI list production, bibliography and catalogue generation, with on-line interactive searching and data entry. The system, shown schematically in figure 2, is based on a Texas Instruments 980A minicomputer with 24K words of mainstore, two magnetic tape drives, two disk drives, a VDU, printer and control teletype. With the two disks the data base can hold up to 200,000 document references. This system was designed and implemented in-house and has been sufficiently successful that it is now being marketed as a general system.

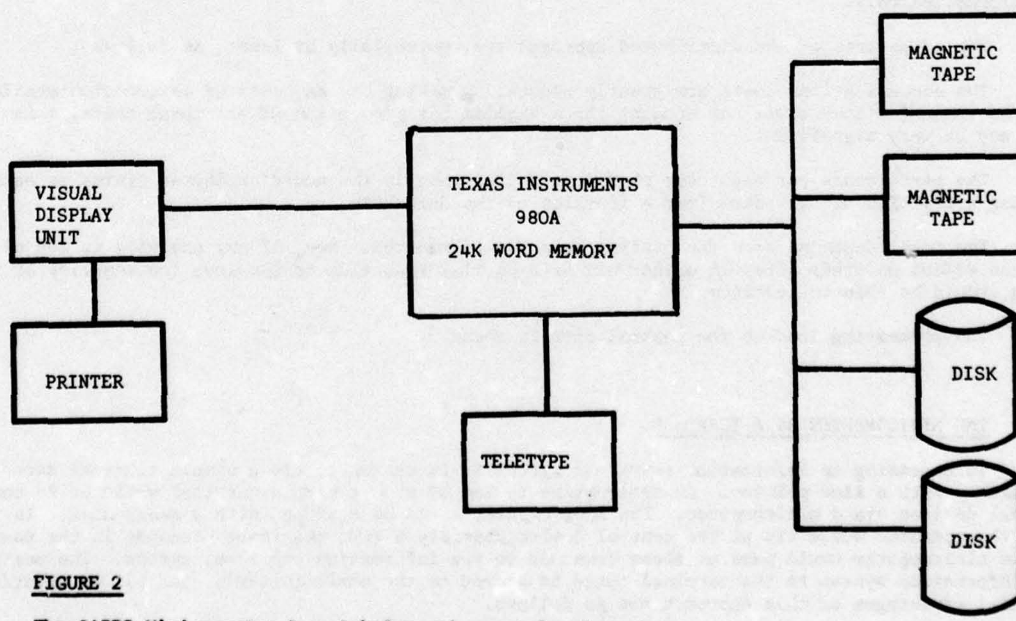


FIGURE 2

The CAIRS Minicomputer based information retrieval system.

7. Distributed Systems:

The distributed information system has the following principles -

a) There are several computers in the total system, all connected in some way to each other. Data is stored at the most appropriate geographical location according to measures such as most common use or cheapness of storage.

b) In operation most requests at a particular site, whether generated on-line or off-line, would be satisfied by reference to the local data base.

c) That minority of requests which could not be satisfied locally would be met by issuing requests to the remote computers.

One particular configuration of a distributed system is the hierarchical system shown schematically in figure 3.

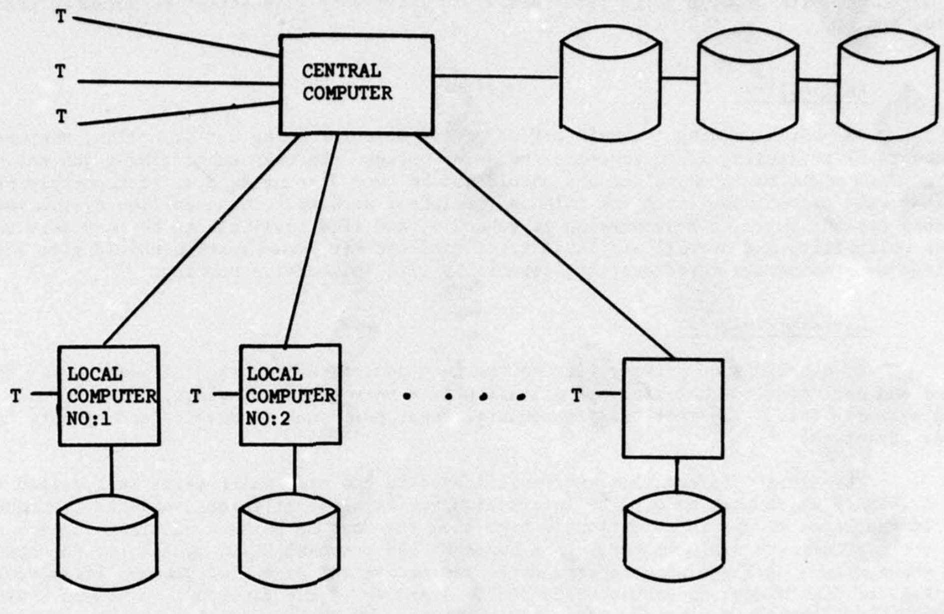


FIGURE 3

Hierarchical distributed system.

Here, each of the remote sites could be based on a dedicated minicomputer, each holding information pertinent to local needs; for example, stock lists of car doors at one site, of engines at another, and of steering assemblies at another. The central site could hold a complete set of all information or summary information only.

The advantages of the distributed approach are, potentially at least, as follows :

a) The communications costs are greatly reduced by making the majority of information available locally. As communications costs can account for a significant proportion of the total costs, this reduction may be very significant.

b) The performance for each user should be better than in the non-distributed system as each site is supplying a fraction of the users from a fraction of the data base.

c) The total dependence on the central computer is reduced. Now, if any one site is out of service, the effect on other sites is either nil or a partial reduction of service; the majority of the processing should be able to continue.

d) The processing load on the central site is reduced.

VI. THE MINICOMPUTER AS A TERMINAL.

In accessing an information retrieval system it is common to use a simple terminal such as a VDU, possibly with a slow printer. An alternative to use of such a simple terminal would be to connect the terminal devices via a minicomputer. The minicomputer could be equipped with a small disk. In operation the operator would sit at the control device, possibly a VDU, and issue commands in the usual way and the minicomputer would pass on these commands to the information retrieval system. The messages from the information system to the terminal could be stored on the minicomputer's disk before printing. The potential advantages of this approach are as follows.

a) Many information systems are accessible to any particular user for a limited period of each day. For example, the 'window' for UK users of the System Development Corporation system is about half of normal working hours. If information can be transmitted rapidly to the minicomputer's disk it could then be printed later, off-line, thereby freeing the available on-line time for more productive work.

b) The costs may reduce as printing which is to be done at a user site will not need to be done during expensive on-line time. In the UK use of information systems costs up to £100 per hour, so any reduction in on-line time will be quite significant.

c) The possibility of doing more printing at users' sites should eliminate postal delays for printing which would otherwise be performed at the information retrieval system and sent on through the posts.

d) If the results of a search have been stored, then the user can add the results to a personal data file. This file can be built up from several sources and as the result of many searches, and could be processed further in many ways. This processing could, for example, be further selection, report generation or even computation on numerical data.

These personal files could be copied onto cassette tape or floppy disk and stored securely and privately in users own offices.

e) The minicomputer could support use of several terminals concurrently, both for accessing its own files ('off-line' mode) and for accessing a remote information retrieval system ('on-line' mode). By appropriate software this on-line use could be made to appear as being simply more intensive use of a single terminal.

f) Not all networks and computing systems allow use of compatible terminals. An earlier paper in this conference (11) described how in Sweden it was necessary to keep three different terminals at one site in order to access three different systems. If access to networks or to computer systems is made via a minicomputer, then the minicomputer, by supporting the required protocol (implemented in software), could allow access to different computers and networks via a single set of hardware (the minicomputer and terminals).

In order to evaluate this technique an experiment has been initiated in the University of London by the Central Information Services unit and Imperial College. An intelligent terminal (an Incoterm) will be sited in the University library in order to provide access to remote data bases. This will have a floppy disk unit attached. The disk will not be large enough to store all data required, nor will the processor or main store be sufficiently powerful to do all the processing required. Instead, the terminal will be connected via a dial up telephone line to a GEC 4080 midi computer at Imperial College, as shown in figure 4. This computer will supply the bulk disk storage and processing power required. It is intended that this system should be in use for two years starting in mid 1977. The Incoterm terminal will be used full time in normal working hours, but the GEC 4080, which is used for other applications, will be used for this purpose for only one or two hours per day.

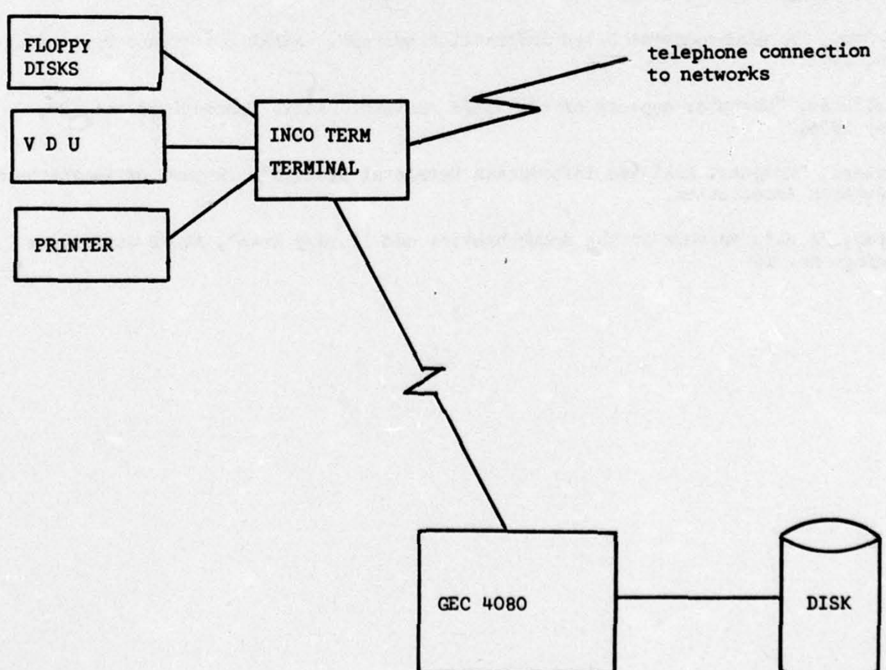


FIGURE 4

Connection of intelligent terminal to remote minicomputer for evaluation experiments.

VII. CONCLUSIONS.

The minicomputer is a viable and desirable alternative to a mainframe computer in several (but not all) aspects of information retrieval systems.

a) As network processors in the communications subsystem of any computer network their use is already widespread.

b) Where the information retrieval requirements are small enough that not all of the resources of a mainframe computer are needed, there are several circumstances where a dedicated minicomputer is superior to a share of a mainframe system in terms of cost, performance and reliability.

c) For interactive on line data entry the minicomputer will give better performance than a general purpose time sharing system, and if the number of connect hours is high enough a dedicated minicomputer can be cheaper than part use of a mainframe.

d) In the replacement, or 'front ending', of simple terminals connected to remote information retrieval systems, they offer several advantages, both economic and functional.

REFERENCES

1. ARPA R.E. Kahn, "Resource Sharing Computer Communications Networks". Proc. IEEE, Nov. 1972.

2. F.E. Heart, R.E. Kahn, S.M. Ornstein, W.R. Crowther, D.C. Walden, "The Interface Message Processor for the ARPA Computer Network". AFIPS Proceedings, Vol. 36, Spring Joint Computer Conference 1970.

3. F. Poncet, J.B. Tucker, "The design of the packet switched network for the EIN project", in Proceedings of the European Computing Conference on Communications Networks, September 1975.

4. N.A. Teicholtz, "Digital Network Architecture", Digital Equipment Corporation.

5. G. Hart, "The use of a minicomputer at the Defence Research Information Centre (D.R.I.C.)". AGARD Conference. Advancements in Retrieval Technology as related to Information Systems, October 1976.

6. R.E. Bleir & A.H. Vorhaus, "File organisation in the SDC time-shared data management system (TSMS)", Information Processing 68, 1969.

7. N.E.C. Isotta, "Space Documentation Service of the European Space Research Organisation". ASIS Proceedings Vol. 9, 1972.

8. R.A. McIvor, "A mini-computer based information system". AGARD Conference Proceedings No. 136, 1973.

9. L.S. Vallance, "Computer Aspects of the BLIPS system". Aslib Proceedings Vol. 27, No.5 May 1975.

10. R. Saunders, "Computer Assisted Information Retrieval Systems". Report of Leatherhead Food Research Association.

11. R. Andren, "A data network in the documentation and library area", AGARD Conference Proceedings No. 158.

THE USE OF A MINI-COMPUTER AT THE
DEFENCE RESEARCH INFORMATION CENTRE (DRIC)

By
GEORGE W HART, B.Sc, DIC
Principal Scientific Officer
Defence Research Information Centre
St Mary Cray
Orpington
Kent BR5 3RE England

SUMMARY

After outlining the functions of DRIC, this paper describes the use of a mini-computer to prepare an Abstracts Bulletin and its indexes, replacing an earlier partly-mechanised system, and to provide data on the exchange of reports with foreign countries. Some of the problems encountered during implementation of the computer systems are discussed. Future possible applications for the computer are described. They include a register of the interests of DRIC's customers, a loans control system particularly for classified reports, thesaurus look-up to help the scientific staff, and information retrieval (both SDI and retrospective). Finally, there is a brief summary of other uses of computers in the UK Ministry of Defence information and library services.

INTRODUCTION

1. During the past year DRIC has obtained a mini-computer which is now used to prepare printing masters for a twice-monthly *Abstracts Bulletin and its indexes*. It also provides quantitative information on the exchange of reports with foreign countries. Later, it will be used to maintain a field of interest register of DRIC's customers, to provide loans information for classified reports, printing recall notices as necessary, and to provide aids for the subject analysts and the editors of the Abstracts Bulletin.

2. It might help to set the scene by first describing briefly the origin and functions of DRIC. The Defence Research Information Centre (DRIC) is the central facility of the Ministry of Defence for the acquisition, announcement and distribution of Defence-orientated unpublished scientific and technical reports originating in the UK and overseas and more particularly those reports which are Defence-controlled or classified. DRIC was formed in 1971 by a merger of the defence element of the former Ministry of Technology Reports Centre and the Naval Scientific and Technical Information Centre (NSTIC) of the Ministry of Defence Navy Department. The Mintech Reports Centre's functions were described in a paper presented by Mr S.C. Schuler to the TIP Specialists Meeting in Ottawa, September 1969 (ref.1, Sec.2). NSTIC had a similar role with regard to report-handling and also acted as a central library for books and periodicals and ran a translations service for the Navy Department. The library function has been taken over by the Ministry of Defence Library Service, but the translations service is still the responsibility of DRIC. The non-defence element of the Mintech Reports Centre now forms the Technology Reports Centre (TRC) of the Department of Industry. The changes are shown diagrammatically in Fig. 1 (on following page). DRIC and TRC may be considered as approximately equivalent to DDC and NTIS, respectively, in the USA.

3. DRIC provides information services to all branches of the Ministry of Defence, their contractors and associates, on Defence Classified and Controlled reports, as well as certain other ancillary services. DRIC has a special responsibility for aerospace information and is the UK National Distribution Centre for AGARD documents. DRIC's staff number 92, including 19 scientists and 6 ADP specialists.

---

1. Schuler, S.C. Experience with a pilot scheme and the transition stage to full mechanisation
In AGARD Conference Proceedings No. 57 - "Problems in Mechanisation of small information centres", September 1969.

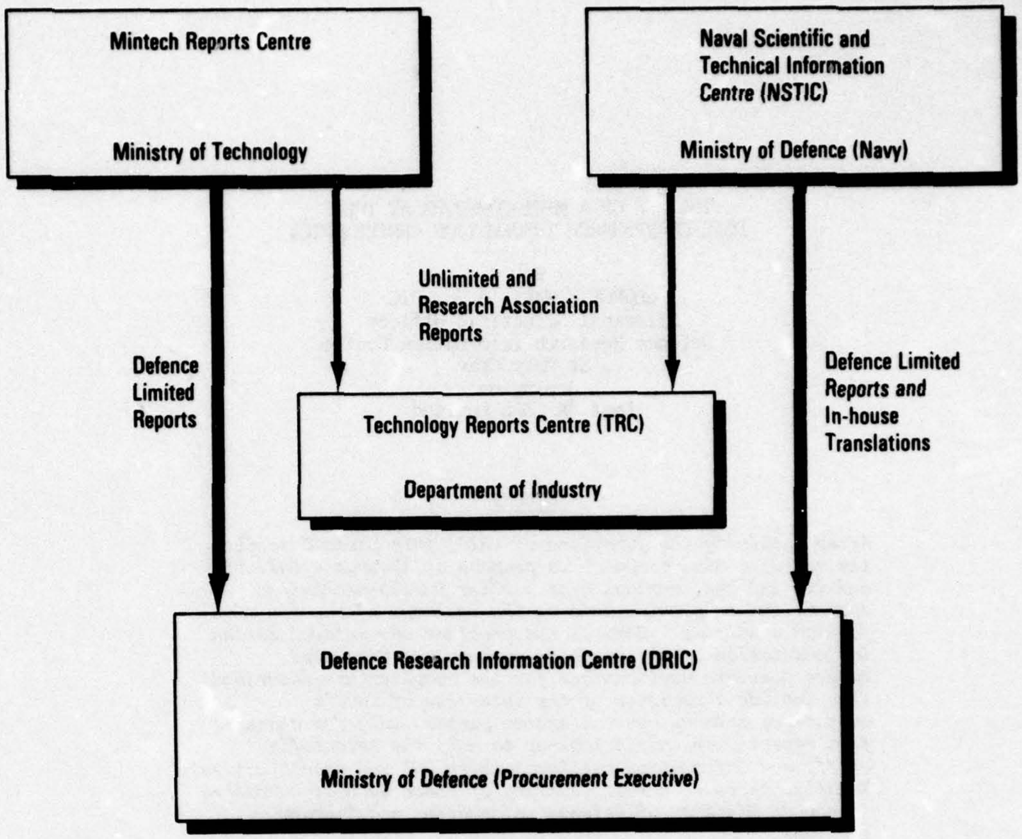


Fig.1. The Formation of DRIC

SERVICES PROVIDED BY DRIC

4. DRIC provides 5 main services and a number of ancillary ones:

(a) Master Record Centre for British Defence Reports

The classification and Conditions of Release of British Reports are checked with the appropriate authorities in MOD and procedures are used to ensure that stipulated limitations on release are observed.

Clearance with Patents Branch and releasing authorities is sought as appropriate. Classified and other controlled reports are transmitted to and from overseas countries via secure channels.

(b) Abstracting and subject indexing UK and foreign Defence reports and publishing a twice-monthly Abstracts Bulletin (Defence Research Abstracts (DRA))

Different editions of DRA are published for MOD and Defence Contractors. Each issue contains subject, author and report number indexes, and DRIC also publishes quarterly and annual cumulative indexes containing additionally indexes of titles, corporate authors, conference papers, translations and contract numbers. A five-year cumulative index on microfiche prepared by Computer Output to Microform (COM) will be published later this year. Magnetic tapes corresponding to the data in DRA in both serial and inverted file form will shortly be supplied each month to the larger R&D establishments to enable them to run their own Selective Dissemination of Information (SDI) services. The software for this will be provided by DRIC.

(c) Distributing these reports to the Defence Community within UK and (in the case of UK-originated reports) to Defence organisations in other countries

British reports are distributed according to recommendations by the originators and the Technical Programme Authority responsible for the work; foreign reports are distributed according to any instructions accompanying the reports. However, DRIC often adds other recipients within MOD in the light of the subject matter and the known interest of potential recipients, depending on the 'conditions of release'.

(d) Responding to requests from the Defence community for such reports

Requests may be met by loans, by retention copies, or by microfiche (always sent for retention), depending on the needs of the requester and the number of copies available. DRIC also attempts to obtain reports, both British and foreign, which have been requested but have not previously been received by DRIC.

(e) Searching the report literature in response to requests for information by members of MOD or by Defence Contractors. DRIC has an on-line VDU terminal to the European Space Agency's RECON system, which holds NASA/STAR and NTIS/GRA files, and will shortly have a similar terminal for searching, back to 1970, data from DRA which are held in computer-readable form at a MOD computer bureau. Manual searches of DRA are carried out at present, and will still be needed for items earlier than 1970 and other data bases. DRIC also provides a limited monthly SDI service for members of MOD, based on the reports listed in DRA. Through the agency of the Atomic Weapons Research Establishment, Aldermaston, a similar service is provided from several of the major data bases such as NTIS/GRA, INSPEC and Chemical Abstracts Condensates.

(f) The ancillary services include translating non-English language documents for R&D establishments and supplying foreign military specifications. DRIC holds a complete and regularly updated set of US Military Specifications and related documents on microfilm (about 105,000 documents) together with hard copies of many of them (mainly defence and aerospace items).

The in-house computer is currently used for items a & b above, and may be used for item e later.

5. The main flows of information within DRIC are shown in Fig.2, which includes some statistics relating to operations during 1975. Thus some 13,000 report titles were handled, 5000 from the UK and 8000 from abroad, and about 110,000 copies of reports were distributed. Defence Research Abstracts contained 9000 items. Requests for reports numbered about 40,000; for military specifications about 8000; and for literature searches and translations, about 600 each.

USE OF THE MINI-COMPUTER FOR THE PRODUCTION OF DEFENCE RESEARCH ABSTRACTS (DRA)

6. The mini-computer at DRIC (a British GEC 4080) which was commissioned in 1975 is shown in Fig. 3, and its main features are listed in Fig. 4. It should perhaps more properly be called a midi-computer as it has more memory and larger discs than most minis. It is a 32-bit word machine with (at DRIC) 64K bytes of semi-conductor store and 19.2 megabytes on 2 fixed and 2 exchangeable discs. Up to 256K bytes of store are available with one CPU, and GEC plan in future to support at least 4 x 70 megabytes of disc storage. The most interesting feature of the computer is the 'Nucleus', a hardware micro-programmed executive, which manages store protection, communications between programs, and short-term scheduling of programs in the system. This enables the computer to provide extensive operating system facilities which would otherwise be impractical because of high overheads. The Disc Operating System (DOS) includes extremely powerful Data Management software which provides device-independent data transmission to and from the peripherals for both applications and operating system programs. The program does not need to specify which files or devices are to be used, this being done at run-time by a command which selects files and peripherals for connection to the 'streams' (essentially input and output ports) of a program. Up to 256 programs or processes may be handled concurrently. DRIC's installation has 3 VDUs and an upper and lower case line-printer. Languages currently supported are FORTRAN IV, CORAL 66, and GEC's own language, Babbage - an assembly language written in the form of a high-level language and in consequence much easier to use than most assembly languages. Interactive Basic and Algol 60 will also be supported under a new operating system recently announced by GEC. Other organisations such as universities have written compilers for other languages.

7. To understand how this computer is used for the production of DRA it might be helpful first to summarise the way in which DRA was prepared until the end of 1975. This is shown diagrammatically in Fig.5a (p2-6) and is described in Section 4 of Ref.1 and in more detail in Ref.2. These references both describe a system using a different make of tape typewriter, but only minor changes had been made by 1975.

8. Reports received by DRIC were indexed for subject content by the scientific staff who also prepared abstracts. After technical editing and the addition of COSATI subject codes, the indexing descriptors and abstracts, together with the bibliographic information (but not the COSATI codes), were punched on paper tape using Datamatic tape typewriters. Details of Corporate Authors were inserted at this stage from edge-punched cards (with an 8-hole paper tape format) selected manually. Two typed outputs were produced. The first, on a so-called 'flimsy', gave a complete listing of what was punched and was placed in an envelope with the corresponding length of paper tape. The second output, a card printed through a carbon patch on the flimsy and containing only the bibliographic details, was used by the appropriate report-handling section of DRIC as a document 'movements' card on which the number of copies held, any initial distribution, any loans and returns were recorded.

---

2. Ridler, E.H. et al — Production of R&D Abstracts and printed indexes to R&D Abstracts using tape typewriters

TIL Report 21, Aug. 1969

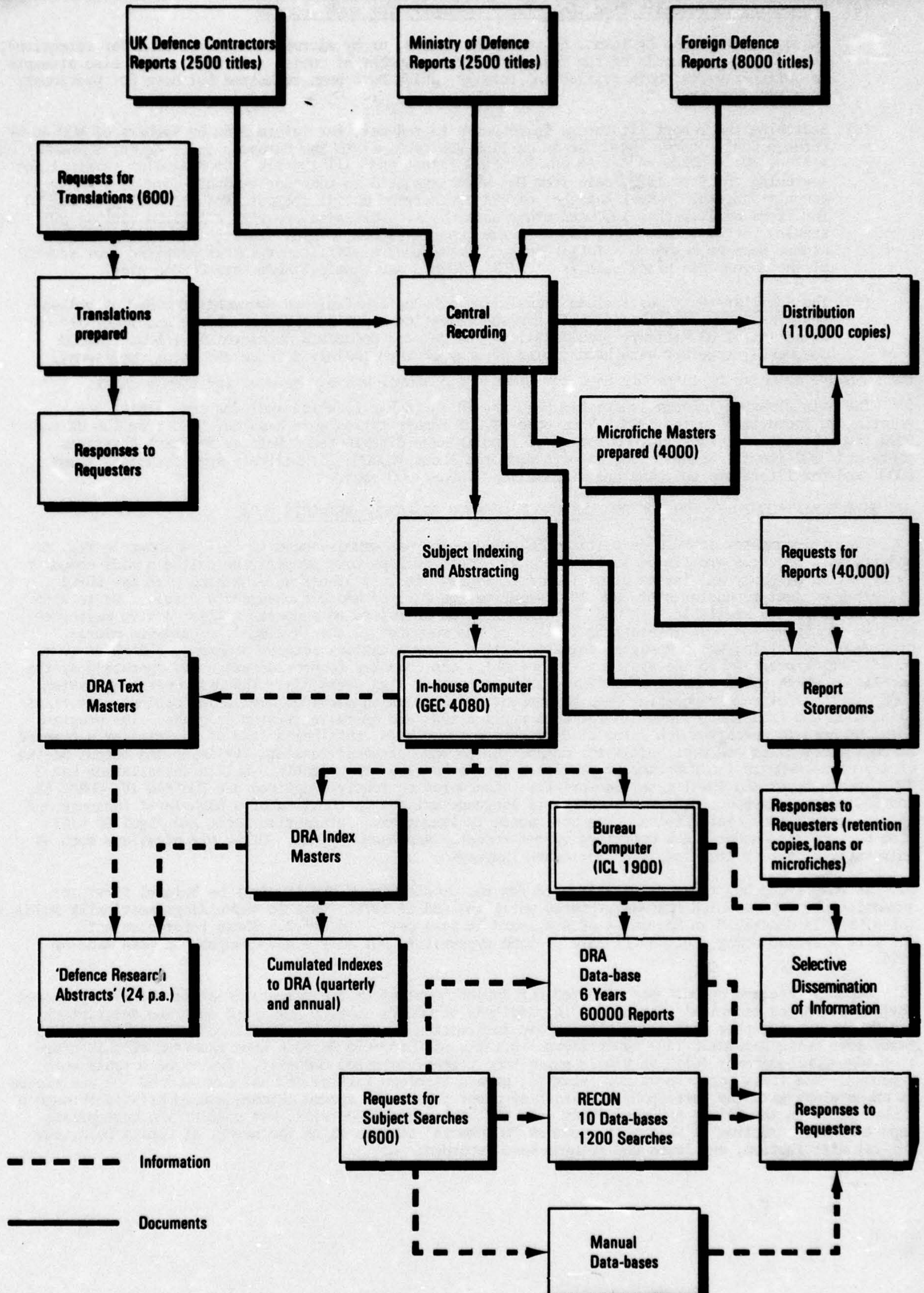
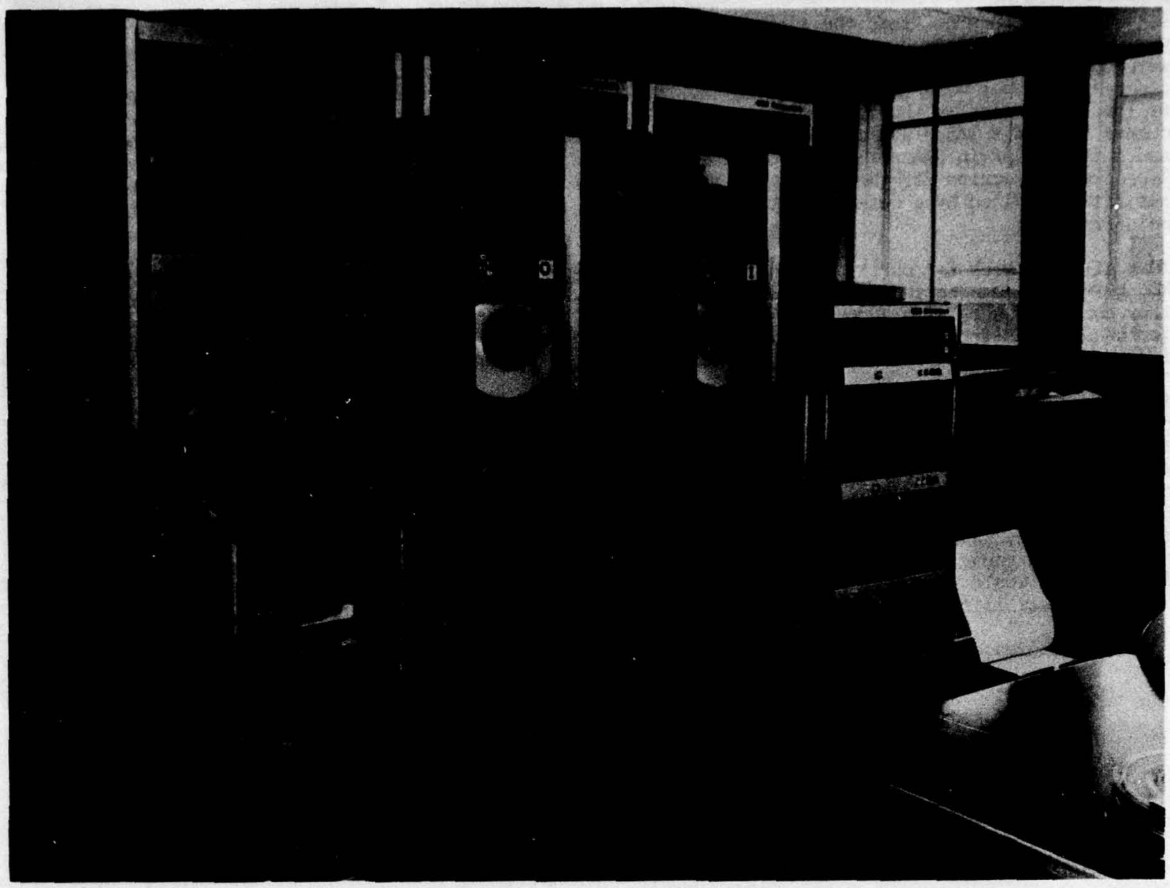


Fig.2 Main Flows of Documents and Information within DRIC

(Figures in brackets indicate current annual totals)



CPU
Paper tape reader
Paper tape punch
Central store

Magnetic tapes

Discs

Line-printer

Fig.3 DRIC's GEC 4080 Mini-Computer

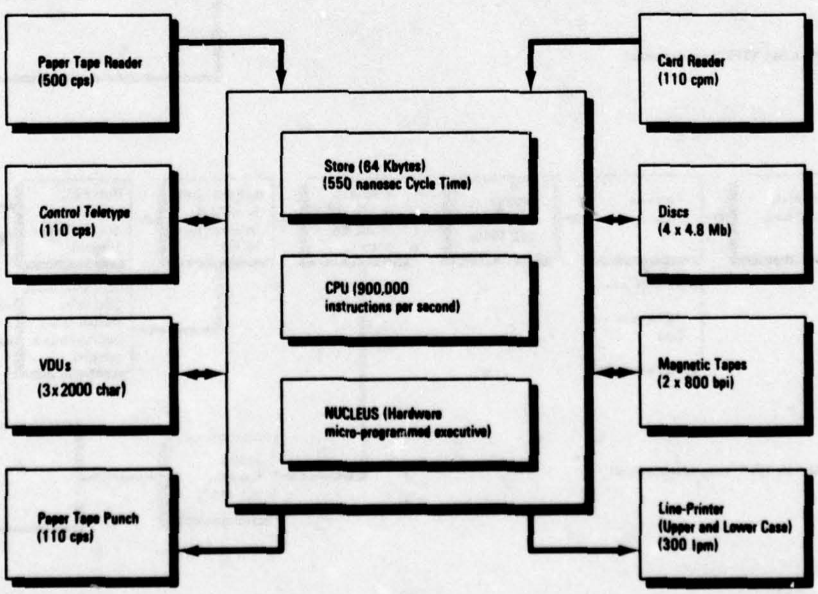


Fig.4 Configuration of the GEC 4080 Computer at DRIC

9. Clerks then proof-read the 'flimsy' against the original document, checking for typing errors, and marked any errors in red. The envelopes (containing the annotated flimsies and corresponding tapes) were then sorted manually by the COSATI codes, this being the order chosen for presentation in DRA. The ordered envelopes were then returned to the tape typists who used the punched tapes to operate the typewriters automatically thus producing a second tape, stopping as necessary to make any corrections. The output from this stage was a number of paper tapes, each containing corrected details of some 25 reports, and a typescript in DRA format. The latter, which was larger than the A4 size of DRA, was sent to DRIC's Reprographic Section for the production of DRA copies by offset litho. The Contractors' Edition of DRA was later compiled by a cut-and-paste method from the masters used for the MOD Edition, unsuitable items being removed.

10. After a check for parity errors, the long lengths of tape were sent to an ICL 1900 computer at an MOD bureau for production of the indexes to the MOD edition of DRA and for the SDI service. The data they contained were also added to a data base held on magnetic tapes at the bureau and covering items included in DRA from 1970. Because the Contractors' Edition was prepared by a cut-and-paste method, no indexes could be provided for it by the computer.

11. It was felt that this system made too many demands on the clerical and typing effort and that it could be simplified by using a mini-computer, with an upper and lower case line printer. Moreover, the mini would have other advantages: it would more readily allow suitable items to be printed in DRA under more than one COSATI subject heading; it could produce indexes which were more legible than the upper case indexes prepared on the 1900 computer; and the masters for both the MOD and the Contractors' Editions could be prepared directly from the computer, allowing the latter Edition to have indexes for the first time. And, since the work would be done at DRIC, and the computer would be under our own control, turn-round would be quicker and it would be easier to control the quality of the print-out. The computer would also have other applications in DRIC (discussed in paragraphs 18-25 below). A detailed feasibility study concluded that the installation of a mini-computer in DRIC would:

(a) increase efficiency in the preparation of data for DRA and indexes;

(b) increase security in the production of DRA (by reducing the amount of movement of classified information);

(c) increase security in the movement of classified documents (see paragraph 22 below); and

(d) provide a generally improved service to all DRIC customers.

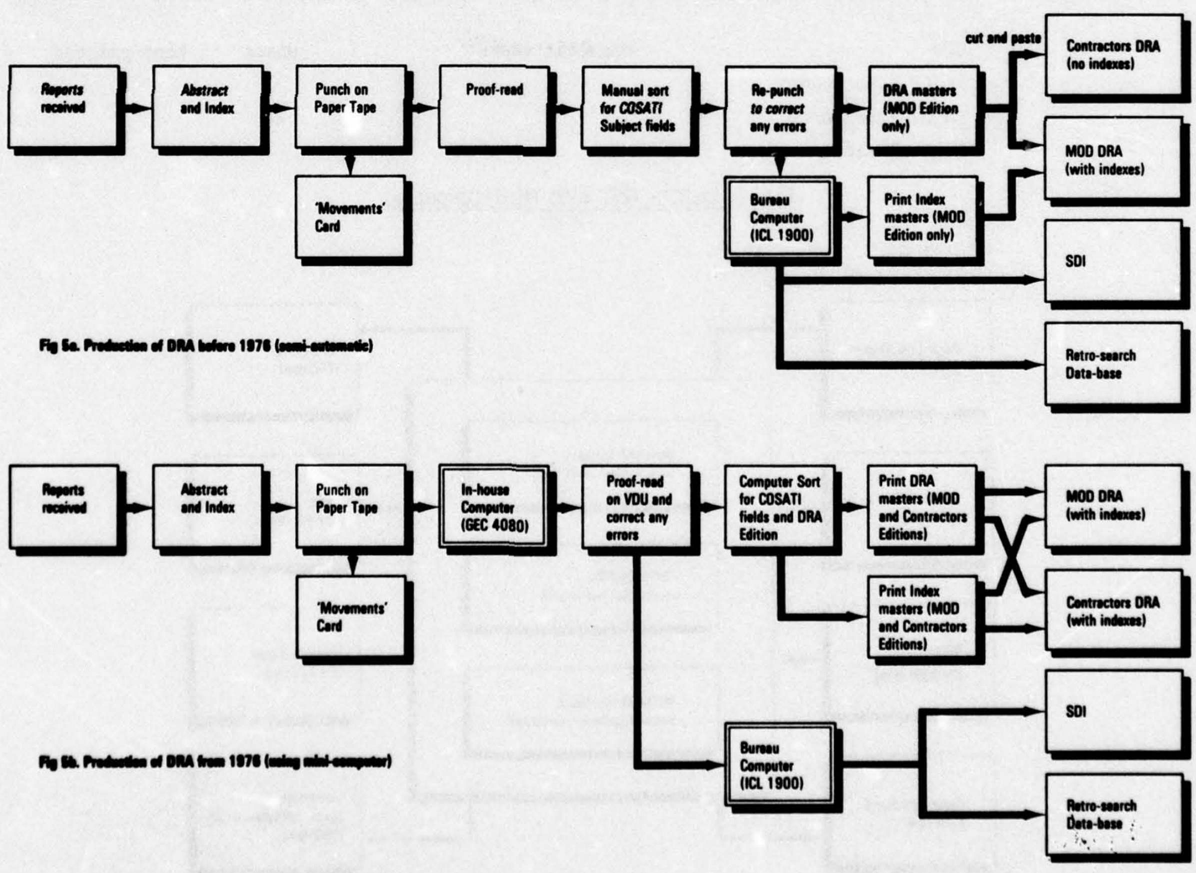


Fig.5 Production of DRA - (a) before 1976; (b) from 1976

12. The system finally devised for the production of DRA is shown in Fig.5b. The first 3 stages are essentially the same as for the previous manual system (Fig.5a), including the production of the movements card, the main difference being the addition of data fields to the punched tape to indicate the COSATI heading(s) and the edition of DRA. However, many items are now punched in random order onto one tape and this tape is read straight into the 4080 computer. Here a number of validation checks are applied, such as checking that the accession number is of the correct form, i.e. has the appropriate number of letters and digits correctly arranged, that an acceptable security classification is present, that the date field is correctly constituted and represents an acceptable date, and that the number of data fields is correct.

13. The next stage will be to display the complete details on a VDU screen together with an indication of any errors identified during the validation run, as shown in Fig. 6. Errors shown in this figure are a date with the month numbered '15', and a descriptor longer than 40 characters(a '/' having been omitted). The errors are in fact shown in reverse video (white on black) to make them stand out more clearly, but this could not be shown in the figure. The proof checkers will work directly from this display and correct the errors indicated by the computer and any others they notice by reference to the original report, such as the mis-spelling of 'gusts' and 'stratosphere' in Fig.6, using the keyboard editing controls. To save time only a sample check of abstracts (as opposed to bibliographic details and indexing descriptors) is made, a test having shown that few significant errors arise there. The second punching of paper tape is rendered unnecessary by this use of the VDUs. We intend to investigate the possibility of direct keying to the computer at a later date, which would simplify the operation still further and allow validation checks to be made at the data preparation stage. Other possibilities are key to magnetic tape or disc.

```
      01-01
      BR-19224 ?A?ARC-CP-1091? ?
      UNLIMITED?
      027500Aerodynamics Dept.,Royal Aircraft Est.,Bedford,U.K.?
      Atmospheric Gusts-A Review of the Results of some Recent
      R.A.E. Research?
      ?
      ?
DATE  Burnham,J.? ?15.1976?59pp 57ref?  ?
      ?  ?  ?
      ?  ?  ?

      *Wind(meteorology),Gusts,Aircraft/Aircraft,Gusts,Thunderstorms
>40c  /Stratosphere/Flight characteristicsMathematical models/?
      Recent RAE research on guests has been particularly concerned
      with severe gusts and the situations in which they occur. In
      the sratosphere mountain wave conditions and those in the
      vicinity of thunderstorm tops have been investigated. At lower
      altitudes, gusts in and near thunderstorms have also been
      studied,as have wind and gust effects likely to be significant
      during take-off and landing. This work has relevance both to
      aircraft operations and to aircraft design. In the latter
      connection, recent work on mathematical models of severe gusts
      is also described. Mention is made is made of the effects of
      pilot control activity during flight through gusts.   PBP?
```

<u>Fig.6. Example of the Display on a VDU for Proof-checking</u>

(The question marks identify the ends of data fields, some of which, eg contract number and similar details, are not applicable to this example and have been left blank)

14. The corrected items for a whole period (half a month) are sorted by the computer into COSATI subject order and according to the appropriate editions of DRA, being duplicated by the computer if they appear in both; and cross-reference entries without an abstract are generated for subsidiary COSATI headings. This sorting of about 400 envelopes used to take a considerable time when performed manually. The masters for printing both editions of DRA, including the indexes, are produced on the line printer. The opportunity was taken to improve the lay-out of the indexes and examples of the old and new subject indexes are shown in Fig.7. A magnetic tape is also prepared by the mini-computer for use in building up the data base and running the SDI service on the ICL 1900 at the MOD Bureau.

7 SUBJECT INDEX 7

SUBJECT HEADING: TITLE	DRIC ACCESS NO.	LOCATION
COMPUTER AIDED DESIGN,PRINTED CIRCUITS COMPUTER AIDED DESIGN OF PRINTED CIRCUIT BOARDS	BR-49112	1091-7603
COMPUTER PROGRAMMING,OPERATING SYSTEMS(COMPUTERS),COMPUTER PROGRAMS PROGRAM DEVELOPMENT OPERATING SYSTEM FOR THE ARGUS 500L COMPUTER	BR-49502	1082-7603
COMPUTERIZED SIMULATION,INDUSTRIAL PLANTS,DECISION MAKING QUARTERLY BULLETIN OF THE DIVISION OF MECHANICAL ENGINEERING AND THE NATIONAL AER	P-209248	1003-7603
COMPUTERS,AUTOMATIC CONTROL,AIR TRAFFIC CONTROL USE OF COMPUTERS IN AIR TRAFFIC CONTROL	P-208860	1239-7603
COMPUTERS,BIBLIOGRAPHIES OPERATIONAL COMPUTER-BASED TECHNOLOGY REFERENCE LIST	BR-49211	1079-7603

Fig.7a Old Form of Subject Index to DRA (prepared by bureau computer)

7 SUBJECT INDEX 7

Subject Heading & Title of Report	Access No	Location
COMPUTER AIDED DESIGN PRINTED CIRCUITS Computer Aided Design of Printed Circuit Boards	BR-49112	1091-7603
COMPUTER PROGRAMMING OPERATING SYSTEMS(COMPUTERS) Program Development Operating System for the ARGUS 500L Computer	BR-49502	1082-7603
COMPUTERIZED SIMULATION DECISION MAKING Quarterly Bulletin of the Division of Mechanical Engineering and the National Aeronautical	P-209248	1003-7603
COMPUTERS AIR TRAFFIC CONTROL Use of Computers in Air Traffic Control	P-208860	1239-7603
BIBLIOGRAPHIES Operational Computer-based Technology Reference List	BR-49211	1079-7603

Fig.7b Present Form of Subject Index to DRA (prepared by in-house mini-computer)

15. A list of corporate author entries is to be held on the computer and appropriate entries will be added automatically to the bibliographic details, the appropriate reference number being punched at the initial punching stage instead of the full details. The errors inherent in maintaining large numbers of duplicate edge-punched cards will then be eliminated. Because the original 6-digit numbering system for corporate authors had proved inadequate to cope with organisations having large numbers of subdivisions, such as the UK Ministry of Defence, it was decided to revise the list according to standard cataloguing rules (with minor local variations) and assign new numbers, incorporating check digits, to each entry. DRIC had insufficient effort available to do this in-house and so a contract was placed for it to be done externally.

Thesaurus Checking

16. The next stage will be to store the complete Thesaurus on the computer. DRIC uses the Thesaurus of Engineering and Scientific Terms (TEST)(Ref.3)with local additions to meet the demands of recent technology and to cope with identifiers such as MRCA and Chieftain. Storing the Thesaurus on the computer will have two main benefits. Firstly, it will enable the subject indexers and technical editors to make reference to an up-to-date list of added terms, using VDUs. Secondly, it will enable the computer to check the acceptability of the descriptors and the accuracy of punching at the validation stage, any errors being displayed on the VDU. The proof-checkers can correct wrongly-spelled or mis-punched descriptors, but any which are not found in the Thesaurus need to be listed for a senior indexer to check. Acceptable new descriptors will then be added to the Thesaurus for future use.

3. Engineers Joint Council in co-operation with the Department of Defense, USA. Thesaurus of Engineering and Scientific Terms

First Edition, December 1967

17. DRIC holds a copy of the original version of TEST on tape at the bureau computer. This has been used to check past indexing and produce a list of descriptors used which are not in TEST. A contract was recently placed to correct spelling and punching errors and to list the others so that they could be considered by the indexers as potential additions to the Thesaurus. The up-dated Thesaurus will shortly be stored on the mini-computer, for use as described above. On the bureau computer, the indexing terms used in the past will be corrected so that retrospective searching can be carried out using this Thesaurus without the user also needing recourse to lists of other terms used.

OTHER USES OF THE MINI-COMPUTER

Reports Exchange Data

18. Another current application for the in-house computer is the production of statistics on the exchange of reports with foreign countries. DRIC has often been asked, e.g. by members of MOD HQ concerned with international cooperation, to provide statistics on the interchange of reports with foreign countries, and these requests have required lengthy manual counts in the past. Moreover, to simplify the actual preparation of statistics, the clerical staff listing reports sent to or received from abroad have used different forms for each major country. This, of course, created problems when more than one person needed a particular form, and time was lost in changing from one to another. Using the mini-computer to prepare statistics has allowed any number of different countries to be entered in random order on a single form. Thus each clerk now uses only one form no matter which countries he is dealing with, and no one else needs to use it. Minor changes had to be made to the format to make the forms suitable for ADP work, but these changes were introduced to the staff concerned in a short talk which concentrated on the benefits to them individually and to DRIC generally and also emphasised the need for care in completing ADP forms. The change-over has gone reasonably smoothly and the forms are generally being prepared very carefully. So much so, that when trying to demonstrate the input validation program, three months after the change, we were unable to find a recent tape with errors and had to create some artificially! This is a tribute to the accuracy both of the clerical staff completing the forms and of the tape typists punching from them. There are occasional problems with unclear hand-written details of destinations (organisations not countries) but these are less important than the other details recorded since they are merely used as records and not as elements in the statistics.

19. Standard statistics are now produced each quarter showing the number of classified and unclassified reports sent to and received from each country. More specific breakdowns by individual months and specific countries can be obtained as required, and these listings include the DRIC accession numbers and the precise classification of each report to enable us to determine in more detail the nature of the trade with the countries concerned. An example of the quarterly statistics is shown in Fig.8. Fictitious names have been used to avoid disclosing possibly sensitive information.

DEFENCE RESEARCH INFORMATION CENTRE

REPORTS EXCHANGE DATA OCT - DEC 1976

R=REQUEST D=DISTRIBUTION

COUNTRY	UNCLASSIFIED IN R	UNCLASSIFIED IN D	UNCLASSIFIED OUT R	UNCLASSIFIED OUT D	CLASSIFIED IN R	CLASSIFIED IN D	CLASSIFIED OUT R	CLASSIFIED OUT D	TOTAL IN	TOTAL OUT
Brobdingnag	4	33	11	54	3	13	8	18		
	37		65		16		26		53	91
Erewhon	0	1	0	0	0	0	0	0		
	1		0		0		0		1	0
Freedonia	0	1	0	3	0	0	0	2		
	1		3		0		2		1	5
Ruritania	0	0	1	0	0	0	0	0		
	0		1		0		0		0	1

Fig.8 Example of Quarterly Statistics of Reports Exchange

20. At a later date we intend to make the output from the specific listings available on magnetic tape so that the titles or the complete bibliographical details and abstracts can be extracted from the main data base stored at the ICL 1900 computer bureau. This will make identification of the subject areas much easier. This is not as straightforward as it might seem since the data on the 1900 is stored not by accession number but by location reference number, a serial number indicating the position and the issue of DRA in which the report was announced, and the two numbers are totally unrelated.

21. Eventually, we hope to provide VDUs for the input of the reports exchange data, with on-line validation. But this must await a feasibility study to determine both its practicality and whether it is cost-effective.

Security suite

22. Another application for the mini-computer is a system for recording the movement of classified reports. As mentioned in paragraph 12 above, the present system provides a 'movements' card on which details of initial distribution and subsequent requests are recorded. The movements cards are held in accession number order, there being only one for each report. The only other way to obtain report details is from the corporate author (originating body), another card being filed in this order for each report and acting as a cross-reference to the movements card which contains the full details. However, if document movements were entered on the computer, any desired listing could be obtained, and it is our intention to do so if a feasibility study shows that it is practicable. This system will be in batch mode initially, but it is possible that on-line working could be developed for this operation also.

Field of Interest Register

23. When requests for reports or searches are received from organisations outside the Ministry of Defence, DRIC has to check that they have an appropriate 'need to know', ie that they are carrying out work for the Ministry in the subject area concerned. To this end, it is intended to prepare and maintain on the mini-computer a 'field of interest' register listing Defence contractors and the areas of interest for which they are accredited. This will be printed in alphabetic order at regular intervals, but it will also be possible to ask for individual entries to be displayed on the VDUs. Another use for the register will be to maintain a list of the subject interests of MOD customers to help DRIC make an initial distribution of reports.

SDI and retrospective searching

24. Two applications dependent upon considerable expansion of the computer are SDI and retrospective searching. The ICL 1900 computer used for these purposes is likely to be replaced by a 2900 'New Range' machine in a few years' time. Rather than re-program for that machine, we intend to assess the possibility of using the mini-computer.

25. A monthly SDI should be straightforward on the mini-computer, once the retrieval program has been written, but retrospective retrieval from an on-line data base may require more disc storage than will be available even with the enhancements planned by the manufacturers. To overcome this difficulty, we propose to examine the possibility of storing only the inverted files on-line and using a computer-controlled automated microfiche reader to hold the bibliographical details and the abstracts. Such a system was investigated by Project INTREX (ref.4) and has been used by the European Nuclear Documentation Service (5). Commercial systems are now available or under development (6). If we do use such a system, we should also be able to provide an on-line service for a few large MOD establishments, who would each hold a duplicate set of fiche. This would be particularly appealing since the communications system would carry only search statements and the numbers of the documents matching those statements. Thus no additional security problems would arise.

---

4. Overhage, C.F.J. & Reintjes, J.F. — Project INTREX: a general review
Information Storage and Retrieval, Vol.10 No.5/6, May/June 1974, pp 157-188

5. Vernimb, C. — Abstracts on microfiche for on-line retrieval
In AGARD Conference Proceedings No.136 - "New Developments in Storage, Retrieval and Dissemination of Aerospace Information", October 1973

6. Duerden, F. — A computer-controllable ultrafiche terminal
In AGARD Conference Proceedings No. 136 - "New Developments in Storage, Retrieval and Dissemination of Aerospace Information", October 1973

PROBLEMS IN IMPLEMENTATION OF THE MINI-COMPUTER

26. As was only to be expected, a number of problems arose during the early days of using the mini-computer, although none of them was unduly serious. Perhaps the most important - and certainly the most time-consuming - was the need for parallel running. It had been decided at an early stage that parallel running of the old and new systems for producing DRA was necessary but we failed to realise the full implications of this decision until the work was well advanced. For true parallel running it would have been necessary to have punched the data twice over, but this would have required considerable extra effort by the tape typists and was totally impracticable, particularly if it were to be continued for 3 months as seemed likely to be necessary. The work could not have been carried out elsewhere because the typists would have needed special training, some of the codes used by DRIC might not have been found on standard tape typewriters, and it would probably have been difficult to find an MOD installation able to carry out the work or a bureau allowed to handle classified material.

27. The input format for the mini-computer system had been chosen for its ease of use on the mini and differed significantly from that used previously on the ICL 1900 computer. Item separator symbols were placed before the end of a line instead of after it, and two fields had been interchanged to make their order more logical. Moreover, two essential new fields had been added, the COSATI subject headings and the edition of DRA; and the mini-computer code for end of record (CR,LF) differed from that of the ICL 1900 (Newline).

28. To avoid the necessity for double punching, it was decided to punch the data as for the mini-computer and convert it by a special program into the format required for the ICL 1900. Although this was not true parallel running, it enabled the new system to be tested in its final form since any errors in the special program would affect only the old system, which was otherwise unchanged. This did in fact prove to be a satisfactory compromise. The special program had to be written in two parts. Firstly, there was an automatic tape program for the tape typewriters, which removed the new fields and converted the tape into a form acceptable to the ICL 1900 tape-reader. An additional tape-punching was, of course, required but this could be done automatically without operator action, apart from changing tape spools. Secondly a computer program was written for the ICL 1900 to edit the input into a form suitable for the existing index production programs. For this we used the extremely powerful ICL George 3 editing system. In theory, we could have amended the index production program to accept the modified input, but that program had been written under contract 8 years before and had had many modifications since. Not surprisingly, we were very reluctant to touch it, especially as it had caused problems earlier in the year which had taken several months to cure.

29. All this, of course, took a considerable amount of extra effort which had not been planned for, and helped to delay live running by one month beyond the planned starting date of January 1976. This caused another problem. The first two issues of DRA were prepared on the tape typewriter/ICL 1900 system and the remaining four issues for the first quarter of 1976 were prepared by the mini-computer. The indexes for each issue were prepared by the corresponding computers, but the quarterly cumulated indexes could, of course, be provided only by one or the other system, not jointly. To have produced the first quarter's accumulation on the 1900 would have been possible but that would just have deferred the problem until the annual accumulation was needed, because that also had to be produced by one computer and not by both.

30. So it was decided that the first quarterly accumulation must be prepared by the mini-computer. At first, it seemed necessary only to re-punch the first two issues of 1976 some time during the quarter in mini-computer format and then process them in the mini-computer to provide the necessary data for the accumulation. However, it was realised (fortunately) that any errors in the manual sorting for the 1900 (or in the punching of the COSATI headings for the mini-computer), if not detected and corrected, would result in reports being given different location reference numbers (see paragraph 20) by the mini-computer from those they had been assigned manually. This would have rendered the indexes valueless and was considered too great a risk to be justifiable and so another method was sought. Eventually two programs had to be written, one for the 1900 to list on magnetic tape the location reference numbers used together with the corresponding DRIC accession numbers, and one for the mini-computer to accept this tape and change the location reference numbers which it had generated to the values which had actually been used. This, too, took time which had not been planned for.

31. Could these problems have been avoided? Yes, if we had taken a chance and decided not to bother with parallel running (a gamble which would, in fact, have come off, but which it would have been very risky to take) or if we had prepared two separate inputs; and they could have been eased to a considerable extent if we had made fewer changes in the format. Perhaps there is a lesson to be learned that changes should not be made when designing a new system unless their full impact on the old system during parallel running has been assessed and accepted.

32. Other difficulties have occurred in the early stages of running the mini-computer programs but they were mainly caused by infrequent combinations of conditions which had not occurred in the test data, and the only lesson that can be learned from this is the old one that testing must be as exhaustive as possible, but that bugs will still lurk unseen until full production running begins in earnest. Some may never rear their ugly heads - but the probability of that is low in most commercial computing!

CONCLUSION

33. In conclusion, I think it is clear that a small computer such as ours can be used to perform many tasks of a quite sophisticated nature for a relatively low cost - a fraction of the price of a main-frame computer.

OTHER USE OF COMPUTERS IN THE BRITISH MINISTRY OF DEFENCE INFORMATION AND LIBRARY SERVICES

34. Although not strictly relevant to the title of this paper or the subject matter of this session, it was thought that a very brief summary of other uses of computers in information and library work in the Ministry of Defence would be of interest.

DRIC

35. In addition to the GEC 4080 computer, DRIC still makes use of the ICL 1900 computer at a Ministry of Defence Bureau in London. This is used for two purposes: Selective Dissemination of Information (SDI); and retrospective searching. Bibliographic data and abstracts of reports that have been processed on the 4080 computer are sent on magnetic tapes to the Bureau twice a month. An SDI service is provided on these data every month for members of MOD Establishments and Headquarters Branches. These data have been accumulating since 1970 and will shortly be used for on-line interactive retrospective searching by DRIC staff in response to requests from other members of MOD and Defence Contractors. DRIC also supplies magnetic tapes monthly to certain of the larger R&D establishments to enable them to run their own SDI services in-house.

Atomic Weapons Research Establishment, Aldermaston and Foulness (AWRE)

36. AWRE at Aldermaston uses an IBM 370 computer for a number of information and library purposes. One is an SDI service based on published data bases (US Government Reports Announcements, Chemical Abstracts Condensates, INSPEC, Nuclear Science Abstracts, and METADEX). The SDI service was initially provided for members of AWRE staff, but has now been extended to many R&D establishments in the Ministry of Defence, under the aegis of DRIC.

37. The Aldermaston Mechanized Cataloguing and Ordering System (AMCOS) (ref.7) is based on MARC(Machine-Readable Cataloguing) records produced by the British Library. It has two main functions: firstly a selection of references relevant to AWRE is extracted from the MARC tapes using the Dewey Decimal Classification field. This list forms a Potential Requirements File from which the Library staff select books for purchase, and from which required data are extracted to generate the orders. Books are of course selected from other sources as well. Secondly, the MARC records are used for cataloguing purposes, with the addition of local information such as shelf number, and accessions lists and catalogues are prepared by the computer. The catalogue is cumulated quarterly by printing out in COM format on micro-fiche.

38. The Aldermaston Library also runs a semi-mechanized loans system, using punched cards, and produces a UDC/Alphabetical Subject Index List for cross-reference purposes.

39. The AWRE outpost at Foulness has developed an on-line loans control system which has been described in ref. 8.

Defence Operational Analysis Establishment, West Byfleet (DOAE)

40. DOAE has an on-line retrospective information retrieval service based on the reports held by the library. The software for this was developed as an extension to the ICL 1900 computer operating system, GEORGE 3, by the DOAE computer staff. An SDI service from the reports accessions is also run for members of DOAE staff and at least one external establishment. A Library accessions list is prepared by the computer. Records of technical files are also held on the computer, lists being printed out in KWOC format.

Royal Aircraft Establishment, Farnborough (RAE)

41. RAE have three operational Library computer systems: a UDC/Subject Index Listing, a Register of Interest of their scientists, and an Annual Index to RAE's Technical Reports.

National Gas Turbine Establishment, Pyestock (NGTE)

42. NGTE uses the RAE Technical Reports program to produce author and KWOC accession lists of books and reports.

7. Corbett, L. & German, J. — AMCOS Project stage 2: a computer-aided integrated system using BNB MARC literature tapes. Program, Vol 6, No. 1, January 1972, pp 1-33

8. Eunson, B.G. — UPDATE - an on-line loans control system in use in a small research library. Program, Vol 8, No. 2, April 1974, pp 88-101

Royal Signals and Radar Establishment, Malvern (RSRE)

43. RSRE have computerized their Periodicals List and also prepare a KWOC index to their Newsletter, also using the RAE program. The KWOC index will shortly be extended to include report material.

Royal Armaments Research and Development Establishment, Fort Halstead (RARDE)

44. RARDE run an SDI service based on the reports received by their Library.

Other Establishments

45. Other establishments are investigating the possibility of using the computer for on-line retrospective retrieval or for loans control, and several, including some listed above, are planning in-house Selective Dissemination of Information services using tapes supplied by DRIC.

ACKNOWLEDGEMENTS

I should like to acknowledge with much gratitude the help given in the preparation of this paper by my colleagues in DRIC. Any errors are, of course, my own responsibility. Any views expressed are my own and do not necessarily represent those of the Ministry of Defence or HM Government.

REFERENCES

1. Schuler, S.C. — Experience with a pilot scheme and the transition stage to full mechanisation

In AGARD Conference Proceedings No. 57 - "Problems in Mechanisation of small Information Centres", September 1969

2. Ridler, E.H. et al — Production of R&D Abstracts and printed indexes to R&D Abstracts using tape typewriters

TIL Report 21, Aug. 1969

3. Engineers Joint Council in cooperation with the Department of Defense, USA — Thesaurus of Engineering and Scientific Terms

First Edition, December 1967

4. Overhage, C.F.J. & Reintjes, J.F. — Project INTREX: a general review

Information Storage and Retrieval, Vol.10 No.5/6, May/June 1974, pp 157-188

5. Vernimb, C. — Abstracts on microfiche for on-line retrieval

In AGARD Conference Proceedings No.136 - "New Developments in Storage, Retrieval and Dissemination of Aerospace Information", October 1973.

6. Duerden, F. — A computer-controllable ultrafiche terminal

In AGARD Conference Proceedings No.136 - "New Developments in Storage, Retrieval and Dissemination of Aerospace Information", October 1973.

7. Corbett, L. & German, J. — AMCOS Project stage 2: a computer-aided integrated system using BNB MARC literature tapes.

Program, Vol 6, No. 1, January 1972, pp 1-35

8. Bunson, B.G. — UPDATE - an on-line loans control system in use in a small research library

Program, Vol 8, No. 2, April 1974, pp 88-101

MINICOMPUTERS IN LIBRARY CIRCULATION
AND CONTROL

by

Klaus-Dieter Lehmann
Managing Director
Stadt- und Universitätsbibliothek Frankfurt
Bockenheimer Landstraße 134 - 138
6000 Frankfurt am Main
Federal Republic of Germany

SUMMARY

The growing need for information services poses increasing problems for libraries and documentation centers. Data processing techniques provide several possibilities for improvement; among the most recent techniques are the use of minicomputers.

A description of the mode of operation in library loan posting is given, as well as a discussion of the extent and structure of data, linkage possibilities, and special operational features.

For these specialized applications a short survey of the hardware configuration and software of minicomputers is also presented. Comparison is made between a stand alone system and a minicomputer connected to a background computer. This study is based on actual projects existing in the Federal Republic of Germany.

1. BASIC PURPOSE

1.1.

When libraries start with EDP, very often they start with circulation control. This is the situation in Germany in any case with its traditionally structured libraries.
The reasons are:
1. Increase in quantity of the borrowers.
2. Users demand on the library.
3. Shortage of library staff and increasing costs.
4. Inefficiency of manual operation.

The basic for an effective application of computers in the library is given by a large quantity of data and a repeated access to identical data. This takes place in any larger library.

1.2.

The purpose of circulation control is quite simple - to assure that books or other materials the library hold, are available to users who need them, within a reasonable period of time and for a reasonable period of time. The data of circulation control should be gotten as fast as possible and in a cost-effective way.

1.3.

There are three different types of data:
1. Informations concerning circulation data (Date, loan desk ...)
2. Informations concerning the user (Name, address, status ...)
3. Informations concerning the book (Book identification, shelf mark, reminder, classified documents ...)

Libraries of course differ greatly in the importance they attach to one or another of these functions, reflecting differences in the purpose of the libraries themselves. Some have the requirement of very rapid accessibility, others have a lot of special collections with different kinds of using them.

2. LIBRARY CIRCULATION AND CONTROL IN WESTERN GERMANY

Most of the research libraries in the Federal Republic have had no direct connection to computer centers until quite recently, therefore they first developed off-line systems. This procedure has some disadvantages, book reservations are difficult, blocked users cannot be stopped at the loan desk and there is always a gap of information between two processing steps of the computer.

An on-line system is able to give this information at once by direct access. The trend in the last few years in library circulation and control in German research libraries has tended towards the use of on-line systems. At present the following on-line circulation systems are working: Aachen, Bielefeld, Bremen, Dortmund, Düsseldorf, Kiel, Münster.

There have been two alternative possibilities for on-line data processing, to be a participant in a central computer or to use stand alone computer in the library. A new situation is created by the development of package minicomputers to support local library functions, especially in circulation control.

3. HARDWARE OF MINICOMPUTERS

3.1.

The mass production of integrated circuits hastened the development of minicomputers. Today the efficiency of minicomputers is comparable in some fields of application to that of large computers.

Minicomputers normally work with a word length between 8 and 32 bits. The shorter the word length, the larger is the requirement of main storage for retention of data and programs. Usually most of the systems have word lengths of 16 bits. Some of these are adressable by position. The storage is generally a combination of magnetic core storage and semiconductor storage thus building a synthesis of economy and quick access.

Most of the minicomputers have a maximum memory size of 32 K words, several peripheral equipment can be connected to the unit.

3.2.

In addition to the memory size the program structure is also important for the applications in circulation control. The theoretical estimation of the program requirement for circulation control is not easy; as a rule it is safe to assume that most of the program instructions are logical operations (compare, move, displace etc.).

3.3.

Another factor, which effects the storage capacity is the size of operating system. This amount is lost to the user. Normally the operating systems need about 4 K.

3.4.

In addition to the system inherent capacity standards the use of the storage capacity is strongly dependent on the ways and frequency of library managing. It makes a difference whether the computer is a stand alone system or only partly so, whether one or various terminals are connected, whether only check in and check out processes are to be worked off or whether there is a requirement for constant queries about users or books. Moreover there are some criteria for the efficiency of a minicomputer system, such as kinds of input, disc units or data protection; but I want to conclude this theoretical survey, and show you instead two examples of how minicomputers work in research libraries of the Federal Republic of Germany. They are the circulation system of the University Library of Münster and the plans for the future of the circulation system in Hessian libraries.

4. THE UNIVERSITY LIBRARY OF MÜNSTER, A STAND ALONE SYSTEM

4.1.

The University Library of Münster is a traditional closed stack library with more than 1.2 Mio. volumes. The circulating materials are in closed stacks. There are max. 3600 orders for loans per day, on average 2100, actual loans max. 2000, on average 1300, and as many returns. The loan desk are opened 7.5 hours. The system of book call numbers is difficult and heterogeneous.

Shortage of personnel made a computerized circulation system indispensable. Financial considerations and the fact that the university computer was already completely utilized by others led to the decision to install a stand alone computer, type IBM System/3 Model 10. It consists of a central processing unit with a 32 K Byte store, a console typewriter, a line printer, a card unit with various functions, three small discs with 2.47 Mio bytes each and a disc storage with 20.48 Mio. bytes.

The purchase price for the system/3 was about 310 000 $ maintenance charges are about 21 000 $.

4.2.

With a grant from the German Research Society the library embarked on a pilot project to develop a low cost, stand alone circulation system, which should by suitable for use in other libraries with closed stacks.

The systems was designed for the following functions:

a) Instant loan from closed stacks.
b) Complete and quick information on books loaned.
c) Automized check in and check out, reserving books, renewals and reminders.
d) Avoidance of going into the closed stacks by direct passing of orders with only positive response.
e) Perseverance of the call number system (without changes), and no additional book equipment.

4.3.

Each loan is preceeded by a request. User code and book number are entered. The user's identification card (as well as the return entry card) is a 96-column punched card. The book numbers are recorded on optical mark enquiry cards. The card contains a number of columns with 10 positions each. The user inserts the book number in these columns in the form of pencil marks. The multi-purpose card unit can read the punched cards as well as the optical mark enquiry cards.

When a request slip is turned in, the information on the book - request granted, or otherwise, the whereabouts of the book - is returned in less than 30 seconds. The recording procedure itself takes 1.4 seconds. Preparing a book for loan requires a half hour on the average. The reasons for this relatively long time are not to be found in the area of data processing, but rather in the poor personal situation in the stacks.

The low disc capacity makes it necessary to organize the book file as a negative file (transaction record of loaned books). The book file, abbreviated user file, and the user address file are indexed. A good deal of effort had been spent searching through the requests to sort out those books which require special treatment or which are not kept in the closed stacks. This information is contained in auxiliary files on disc stores.

On the whole, the information length is kept very brief, in part even as bit information. The whole software, with the exception of the programs pertaining to the operating system has been newly written in programming language RPG II (Report Program Generator II). The developmental phase required five man-years.

4.4.

The evaluation of the procedure shows the following results:
With practically no increase in personnel, the loan quota rose by 70 %. The circulation of books has been accelerated. The service is quicker than before. The information on the whereabouts of the books is varied and informative. The down time of the minicomputers is insignificant and hardly affect the loan system. Another advantage is the flexible preventive maintenance time, which can be arranged outside of operating hours, and the fact that the system can be put to use specifically according to the prevailing conditions.

It is to our disadvantage that any enlargement of the main memory unit and the addition of further terminals are nearly impossible, to say the least very costly. The connection of several input units in more or less distant locations is hardly possible. The capacity of the printer is also dissatisfactory at 6000 lines per hour.

5. CIRCULATION CONTROL IN HESSIAN LIBRARIES

5.1.

The development of the use of EDP in the Hessian libraries took a different course. Allow me to make some basic observations on this subject in advance. In 1970, five regional community computer centers with identical computer systems where founded in Frankfurt am Main, Gießen, Darmstadt, Kassel and Wiesbaden. They are responsible for the automation of different functions in public administration. The public libraries and research libraries are both linked to this network. The goal is to achieve an integrated library system step by step. For this reason, all processes are being developed under the aspect of universal applicabiblity, allowing the participation of all types of libraries. The processing of local data takes place in each regional computer center; data to which several libraries must have access (bibliographical data as a rule) is processed in central files in the regional computer center in Frankfurt. The data processing network operates free of cost for the state and community administrations. The state of Hessen carries the subsidies. As a result, analysis, programming, and processing are carried out by the network whereas the data compilation and input is organized and financed by the individual library. Under these conditions the development of the automation of circulation control was begun in 1971. Included in the process are local loans and active and passive interlibrary loan both from the closed and from the open stacks. Internal lending between the branches of a library as well as borrowing and returning books at a branch of another library is possible.

5.2.

The system operates completely with numerical information on users, books, and participating libraries. For financial reasons we started with an off-line version. The processing was carried out by each respective regional computer center (currently seven libraries are participating). From the beginning the files have been stored on discs. By now this off-line version has been replaced in some libraries by the installation of minicomputers of the system Nixdorf 720 for the input of alphanumeric data (user and book infomation) and a system Nixdorf 710 for the recording of numerical data (loan and return transactions). The system Nixdorf 710 offers 12 K of its main storage for data storage, the system Nixdorf 710 8 K. This modification - employed in Hessen initially by the Gesamthochschulbibliothek Kassel - makes use of job sharing between the main computer and the minicomputer in the library as follows:

The user data and book data are managed separately in the magnetic disc stores of the computer center after they have been recorded on magnetic tape in the library. The book information file covers the complete circulating holdings of the library. The record has a fixed length of 60 characters and consists of: 1) library identification code, 2) book number, 3) book identification, 4) user number (this field is only occupied when the book is loaned out), 5) date, 6) statistics, 7) subject code.

The user file has a fixed length of 92 bits, containing

the library code	the number of loans
the user code	statistical data
the users' personal data	

The data of 10 000 users take up about 4.5 % of the capacity of a disk pack. The indexing system (ISFM) stores the records and the processing is done by using index sequentiell access method (ISAM).

However, reserving of books, reminding and identifying of blocked users is done with the minicomputer in the library. As the system's connection to the central computer is on-line, immediate informations on user and book data are available during opening hours. All batch processing is also done on the main computer.

This way of using the computer is fairly optimal because the processing of the mass of routine data (posting) advantage is taken of the central computer's capacity and its peripheral equipment, whereas for all additional tasks the direct access to the minicomputer and its possibility of immediate processing are made use of.

5.3.

This is the way the system works:
The items of books are identified by means of bar-coded labels. A light pen scans the bar code and recognizes the number. A check digit facility built into the label, ensures the accuracy of the data. Users are also identified by unique bar-coded labels affixed to user cards.

During check out, for instance, the operator passes the light pen over the bar-coded label of the user card and the book. This automatically records the transaction. The recording of a discharge is accomplished simply by passing the light pen across the label of the book. If there is a hold on a book or if it is overdue the system informs the operator by giving an audio-visual signal. Just as reserved books can be indicated on discharge, users whose numbers have been stored can be intercepted at the issue point. All transactions which take place at the terminals are written on the magnetic tape cassette. It is then necessary to feed this data to the computer for the updating of the user file and book file. After that the records for reserved books, reminders and blocked users are stored and available in the mini-computer. This way of writing the data on magnetic tape cassettes and then updating them on the main computer is to be discarded in the near future and to be replaced by immediate input and erasure of the holds, reminders and blockings.

5.4.

We intend to work with that type of minicomputer that provides sufficient storage capacity for additional software and data by disc store. It is safe to assume that for a library with about 600 000 loans per year a disk capacity of max. 2 Mio. Bytes for the tasks of reserving books, reminding and blocking will be sufficient.

6. CONCLUSION

6.1.

Experience has shown that the mechanization of book circulation on minicomputers can led to cheaper costs and more effectivity. The smaller range of the minicomputer's programs and functioning, easier to control, opens up a new possibility for us librarians: once again we ourselves can solve the greater part of our problems and take into consideration the local needs of libraries.

It has also become evident, that a stand alone system with fixed structures which is not connected to a large computer is soon put under too great strain because of the permanently growing demands on it and the necessity to include other tasks of the library into the process of automation.

6.2.

It seems to me, that the development of powerful flexible and cost-effective minicomputers on one hand and the embracing of libraries to networks with a bibliographic data base on the other hand is a good and hopeful evolution. The connection of minicomputers to a large computer within a network solves specific problems of one library as well as gives the possibility of sharing bibliographical data and library resources.

The combination gives libraries the chance to integrate functions within a library and also build up a library network system. Such a concept would combine the best advantages of both.

7. REFERENCES

1. Arbeitsstelle für Bibliothekstechnik. Automatisierung der Ausleihverbuchung. Berlin 1971. 81 p.

2. Aufbau und Einsatz von Minicomputern. BTS 21 (1975) No. 3. p. 5.

3. Bläsing, Rudolf, Heinz Dross und Johann Schubert.
Anwendungshandbuch für die Automation der Buchausleihen in den Bibliotheken des Landes Hessen. Frankfurt 1973.

4. Datenerfassung und Datenverarbeitung in wissenschaftlichen und kommunalen Bibliotheken Hessens.
Seminarveranstaltung Nixdorf. Paderborn 1976.

5. Götze, Wolfram: Das rechnerunterstützte Ausleihsystem in der Universitätsbibliothek Münster (maschinenschriftlich).
Münster 1975.

6. Köttelwesch, C. und K.-D. Lehmann: Gemeinsame Aufgaben der Bibliotheken in Hessen.
In: Bibliothek u. Buch in Geschichte und Gegenwart. Pullach 1976. p. 236 - 257.

7. Lehmann, K.-D.: Bericht über Planungen für den Einsatz der EDV im Bibliothekswesen des Bundeslandes Hessen.
ABT-Informationen 12 (1973). p. 43 - 49.

8. Mc. Cann, Louise, R. Mc Gee u. R.T. Kimber: Comparison of Computerized Loan Systems in the United States.
Program 7 (1973). No. 1. p. 24 - 37.

9. Minnesota Bio-Medical Mini-Computer Project. Journal of Library Automation 5 (1973) March, p. 66.

10. Pudlatz, H.: Die computerunterstützte Ausleihe der Universitätsbibliothek Münster mit dem IBM-System/3.
IBM-Forum. Böblingen 1973.

The Minicomputers Role in Data Recording for Information Retrieval Purposes and printed Information

by

Horst Zuchel

Zentralstelle für maschinelle Dokumentation (ZMD)

Frankfurt(Main), Germany

Summary

It is wise to look continually for possibilities to reduce the mental part in data recording in order to achieve more efficiency. The present state of minicomputer development offers suitable facilities for improving data recording for information retrieval purposes and for printed information.

An approach on that line is being undertaken in ZMD. The hardware configuration, the developed program packages and also the experience gained will be described.

The creation of data needed for information retrieval and computerized production of printed information is a very sensitive task because of the extended font of characters and varieties of structures. All data recording machines have a limited font of characters. At first glance it would seem impossible to process data with an extended font of characters. Therefore it was necessary to find methods to manage more comprehensive fonts.

One of the possible solutions, developed by ZMD in 1968, is the creation of the so-called protype method. A protype stands in place of the original character which cannot be otherwise represented. Protypes are composed in the following form:

a basic character or basic figure and
a two-digit number.

Protypes have to be marked in the text by a preceding control character, e.g. §.
Examples:

Protypes representing special characters
e.g. á

§a22
→ control character
→ basic character
→ two-digit number

Protypes representing special signs
e.g. %

§364
→ control character
→ basic figure
→ two-digit number

This protype method allows the representation of about 6 000 different characters.

The original data does not contain an adequate structure for processing by computer. Therefore, the data must be structured in a unified computer acceptable format. That normally can be done by categorizing and by using control characters. In the data processing stage the computer can recognize the data elements and the contents of the data from the structure. The real problem is to get the structured data free of errors in order to achieve a high degree of precision in the stage of data processing. Therefore it is very important to check and correct the data during the recording stage and before data processing by computer begins.

ZMD uses paper tape typewriters for data recording. The paper tape stored input data then has to be converted onto a magnetic tape as this is more suitable for checking and correcting by program. This program includes checking procedures of general nature as well as subroutines for special checking purposes in special applications. The program works by getting its orders from parameter cards which are permanently stored in the system. These parameter cards allow the program to meet the requirements of different applications and projects. These parameter cards have a specific format in which each card describes the error checking to be performed on one category.
Error checking is conducted on such points as

- completeness of the categories-scheme
- occurence of certain letters, numbers or special characters in certain prescribed positions
- prescribed number of control characters
- interdependence of categories
- length of a record, category, line or string
- ascending identification numbers or ascending tags.

All incorrect data including a corresponding error message are listed by the program.

The detected errors are corrected by

- replacement
- insertion
- deletion

of records, lines or words.
The correction commands have also to be punched in paper tape and in the same way as the original data they have corrected onto the magnetic tape.

This processing cycle of data recording, converting, error checking and correction has been successfully performed for several years; nevertheless it does include essential disadvantages:

- there are too many working steps at too many places performed by too many machines and persons. That means: time wasting and susceptibility to errors
- there are no facilities for automatic error checking during recording
- there are insufficient facilities for direct correction if errors are detected during data recording
- the workload is too heavy for the ZMD computer system (IBM 370/145 as tyme-shaving system, which runs 24 hours per day and is dedicated to information retrieval and text composition activities).

Therefore ZMD was looking for possibilities to avoid these disadvantages in the data recording stage in order to achieve more efficiency. The present state of minicomputer development seems to offer such improvements. Especially since minicomputer systems are equiped with sufficient disk capacity and a powerful central processing unit, which makes the concept of a project-specific checking program controlled by parameters a reality on

a minicomputer. This means the transformation of the mentioned card images into a generator program which specifies the error checking. In other words the minicomputer can execute the same error checking and correction functions as the bigger IBM system of ZMD.

In order to do this a minicomputer used for data recording purposes has to be equiped with:

- disk or floppy-disk units with sufficient capacity
- one magnetic tape unit compatible with the IBM computer to transfer the disk stored data
- visual display unit (VDU) with
  . sufficient capacity for representing a complete document and
  . a suitable font of characters, which includes characters in upper and lower case (The not directly available special characters and special signs could be represented by protypes)
- printing and hard-copy facilities.

ZMD has rented two minicomputer systems, DATAPOINT 2.200 (GIER Company) and CTM 70 (Computer Technik Müller Company), which fulfill these demands.
The special programs for recording, error checking and correction of data have been developed by ZMD. The programs for DATAPOINT 2.200 are written in the problem-oriented language DATABUS, the programs for CTM 70 are written in ASSEMBLER. The system are operational, but we still need to find the final balance between the user demands and the technical aspects. In particular we have to reduce the response time of the DATAPOINT system.

At this stage the following advantages of data recording by minicomputer are clearly recognizable:

- more simple procedures
  Data recording and immediate correction of indicated errors; data recording, error checking and correction without delay and without any unnecessary intermediate steps. All in all this makes the cycle simple and effective.
- reduction of the main computer workload
  The data leaving the minicomputer is free of errors thus avoiding any further checking and correction steps on the main computer. This reduction allows a more effective use of the IBM system for further information retrieval and text composition activities.
- no data conversion
  Data conversion can be omitted.
- more data security
  More data security will be obtained because of less data transportation.
- permanent availability
  Contrary to the IBM system the minicomputer as stand-alone system is available permanently.
- no special requirements
  No special requirements of minicomputer location have to be considered.

In conclusion, experience has shown that a minicomputer in connection with a visual display unit offers considerable advantages.
First, the program controlled data recording with its extensive error checking procedures and prompting functions leads to data free of structural errors.
Second, errors in contents of data can also be corrected subsequently just as in conversational mode.

The minicomputer is able to master successfully recording, error checking and correction of non-numerical data independent of a large computer system. This makes the minicomputer very interesting for libraries and documentation centers.

References

Zentralstelle für maschinelle Dokumentation (ZMD)
Typen und Protypen für den Fotosatz mit der LINOTRON 505.
2.Aufl. (ZMD-A-22)
Berlin, Köln, Frankfurt(Main): Beuth Verl., 1975, 30 S.
ISBN 3-410-44022-4

Zentralstelle für maschinelle Dokumentation (ZMD)
Maschinengerechte Erfassung von Titelaufnahmen mit Lochstreifenschreibmaschinen.
(ZMD-A-23)
Berlin, Köln, Frankfurt(Main): Beuth-Vertrieb, 1972, 48 S.
ISBN 3-410-44023-2

Schneider, K.
Computer aided Subject Index System for the Life Sciences.
München: Verl. Dokumentation, 1976, 205 S.
ISBN 3-7940-2810-4

Arbeitsstelle für Bibliothekstechnik
Bericht über on-line-Datenerfassung mit IBM 3277 und Siemens Transdata 8152.
ABT-Informationen 14 (1974), S.17-37

Fleischer, D.
Erfassung und Prüfung bibliographischer Daten in der Universitätsbibliothek der Technischen Universität Berlin.
In: Arbeitsstelle für Bibliothekstechnik bei der Staatsbibliothek Preussischer Kulturbesitz: Neue Verfahren für die Dateneingabe und Datenausgabe in Bibliotheken.
München: Verl. Dokumentation, 1973, S.131-142
ISBN 3-7940-3213-6

Agi, M.
Datenerfassung mit Sammelsystemen.
Öffentliche Verwaltung und Datenverarbeitung
(ÖVD) 5(1975)2, S.55-57

Göhring, W.
Übersicht über Architektur und Leistung von Kleinrechnern.
Öffentliche Verwaltung und Datenverarbeitung
(ÖVD) 4(1974)10, S. 469-482

INTERNATIONAL DATA COMMUNICATIONS : PROSPECTS AND PROBLEMS

by D.L.A. Barber B.Sc. (Eng.) C.Eng. M.I.E.E.
Director - European Informatics Network
National Physical Laboratory
Teddington, Middlesex
England

1 Summary

The paper briefly reviews the evolution of international data communications, and then considers future possibilities for development in more detail. Some of the issues are illustrated with reference to the European Informatics Network, and the Euronet Project, and the paper concludes with a discussion of problems at the user level.

2 Introduction

Since the advent of Teleprocessing some fifteen years ago, its impact on society has grown remarkably - some might even say alarmingly - in spite of a number of fundamental inhibiting factors. These include the generally unsuitable architecture both of hardware and of software in large computing systems, and the inadequate facilities that are provided by the world's telecommunications services. This paper concentrates on current approaches to improving telecommunication services for the purpose of data-communications, assuming that the present demand for these services reflects the basic need of people to communicate and to obtain information: a trend that has been apparent since the dawn of structured society.

In all human affairs, the immediate future is largely constrained by the present, and existing ideas and facilities are clung to until pressures for change become irresistible. Teleprocessing began, therefore, with the adaptation of the telephone network to carry data between distributed components of computing systems. But the amelioration of the fundamental mismatch between the internal data rates of computers, the external data rates of people, and the transmission rates of the standard telephone channel has required a variety of developments based on the distribution of logic and storage outside the telephone network. This approach to teleprocessing system design still continues.

New technology is, however, rapidly transforming the telephone network and digital techniques are being introduced on an extensive scale. So, another possible way to design new data networks is to exploit these new techniques, which should provide better facilities for data transmission and switching at marginal cost. On a closer examination this approach proves to be less attractive, at least in the context of the data users requirements. Nevertheless, it is still being pursued by some organisations, and this has led to proposals for a new digital data network using synchronous transmission and switching, and even for a future integrated telephone and data network, based extensively on digital techniques.

A more radical approach is to consider the provision of communications facilities between people and computer systems as an overall systems problem. This permits the adoption of solutions based on new principles that are able to provide a data network with properties best suited to user's requirements. Packet switching has been the result. Private data networks use this technique extensively, and most of the current activity in public data networks is in this area.

Packet switching appeals to computer designers because the features it offers meet their system requirements, while they can readily appreciate the principle of dynamic resource sharing on which it is based. Hitherto, communications engineers have seemed less convinced, but the growing number of proposals for national public packet networks suggests this attitude is changing. It has, however, required the construction of a number of experimental packet networks in various countries to achieve this evolution of opinion, and the differences in design of these networks seemed to make prospects for a world-wide public network remote, at least, until quite recently. These issues are considered later, using the European Informatics Network, and the Euronet Project as examples.

The possibility of limited facilities becoming available for international communication using packet switching now appears quite promising, but even if universal data communications facilities were already freely available, there would still be a major hindrance to the free interaction of data processing systems; namely, their mutual incompatibilities. This emphasises the pressing need for standards at the user level, and the paper concludes with some discussion of this topic.

3 Using The Existing Telephone Network

Apart from a few organisations like railways that have their own communication lines, it has been generally necessary to use the telephone network for carrying data between components of any distributed computing system. This creates some considerable difficulties for the users and also for the telephone network, but, nevertheless, this approach has had to be the basis of all data communications schemes in use today. Fundamentally, it has required the use of modems to convert digital signals to analogue form for passage through the telephone network, and has also required the use of storage and logic outside the network to overcome some of it disadvantages. This approach is still being refined and is likely to continue for some time to come.

The advent of Large Scale Integration (LSI), microprocessors, and the present wide assortment of memory devices has enabled better use to be made of the telephone channel, by designing terminals with enhanced capabilities compared with earlier versions, and these developments have more or less kept pace with users requirements. But although world wide communication is potentially available through the telephone network, fundamental disadvantages still remain that make this approach unacceptable as a long term way of providing data services. However, it is worthwhile noting that the flexibility provided by the use of modern techniques, makes it potentially easy to adapt terminals being developed now to suit the public data networks that will be coming into service in the years ahead.

This adaptability may not apply in the case of some other contemporary network building equipment. For example, there now seems to be a rapid growth in the marketing of concentrators, and other equipment incorporating storage and intelligence, intended for use in new private networks. It seems likely that the commercial life of such equipment will be relatively short, because the proposed public networks promise to provide similar features on a more extensive and favourable basis, and there are bound to be strong pressures by the Authorised Telecommunications Carriers to limit the licensing of future private networks.

4 Using The Future Telephone Network

The telephone network itself is rapidly being transformed by the introduction of techniques based on the use of large scale integrated circuits and this process will accelerate. Digital transmission has been used for a long time, and PCM (pulse code modulation) shorthaul links using time division multiplexing (TDM) techniques are commonplace in many countries. At first, all switching had to be done at the analogue signal level and analogue to digital and digital to analogue conversions had to be made at each exchange. Relatively recently, digital switching has been introduced where channels are switched through tandem exchanges unchanged, using digital logic and temporary storage to delay PCM time slots in different links to bring them into alignment with each other.

Looking ahead the advent of analogue to digital convertors, using LSI, suitable for mounting in telephone instruments would bring the digital network right to the subscribers premises. This is not economical at the present time, but could be so in the next few years. But it is very unlikely to be available on a really world wide scale in the foreseeable future, and this is a strong inhibiting factor in the adaptation of this and similar techniques as a basis for new data networks.

A further example of the use of LSI are the developments now occurring in signalling; it is now possible to use push button dialling with the Strowger step-by-step switching system, by means of integrated circuits which store numbers as they are put in, and then translate them into a sequence of pulses, similar to those produced by the normal dial. However, this kind of adaptation does not necessarily confer the uniformity necessary for other uses of the telephone network, and it may prove difficult to employ such a scheme in conjunction with the voice response computer based interrogation systems, designed for use with the conventional type of rotary dial, or the push button dial with multi-frequency signalling. (Ref 1).

In summary, the exploitation of the digital aspects of the telephone network for data does, initially, seem attractive and likely to provide marginal costs for data traffic. But the long delay before these features become universal and the fact that they do not necessarily confer desirable properties for the data user in a network primarily intended to carry speech traffic, makes their impact on future data communications rather less likely than at first sight.

5 An Integrated Telephone and Data Network

The idea of a universal network that would offer all kinds of services to users in a coherent fashion has been discussed for many years. The problem with this approach is to predict users' future requirements, taking into account the heavy investment and the long time scale necessary to create such a network. However, there have been various proposals for an all-digital data network that might, conceivably, be compatible with a future digital telephone network, so that transmission plant and, hopefully, some switching would be common to both networks.

A number of proposals have been discussed within CCITT (International Telegraph and Telephone Consultative Committee) for digital public networks and a particular scheme has been considered quite seriously for Scandinavia (Ref 2). Unfortunately, there is much pressure to design new digital data networks to be similar to the expected form of the telephone network, even if this makes the data network less suitable for its intended purpose. An example is the controversy that has raged over the number of bits to be included in the time slot for each channel.

To accomodate the eight bit byte common in computer systems, a proposal was put forward for a ten bit time slot, with eight bits for data, one bit for framing, and one bit for status, ie. to distinguish data from signalling information which could be interleaved with the data. This is the so called eight plus two bit scheme. Unfortunately, the TDM telephone channel in the USA is based on the use of eight bits in a time slot with seven bits for digitally encoded speech, and one bit for framing: this fact has given rise to the rival six plus two bit proposal in which data is carried in six bit groups, with framing and signalling bits added to make an eight bit time slot. (Ref 3)

A future digital network based on a ten bit channel could have readily accommodated a seven bit speech channel with the extra advantage of providing channel associated signalling, instead of carrying all signalling in a common channel as is the case with PCM telephony today. However, the arguments put forward by telephony system designers, based on the heavy investment already made in old equipment, seem to have moved the choice in the direction of the eight bit channel, to the detriment of a possible future integrated data and telephone network, which would really meet the needs of data users.

It now seems that the plans of some carriers for the provision of a digital synchronous data network by the early 1980's will be delayed to the 1990's. This may be due to recent announcements about public packet switching networks which, if successful, may cause further delay or, indeed, eventually lead to the cancellation of such plans. This means that an integrated network may never be possible on a world scale and that telephone and data networks will evolve independantly, hopefully to meet the requirements of their users.

6 Packet Switching Networks

Since it was conceived in the early sixties (Refs 4,5), the technique of packet switching for data communications has developed rapidly; this has given rise to experimental networks in various parts of the world, and we are now at the dawn of an era when an increasing number of public packet switching networks will be introduced by various communications carriers.

Until quite recently, many of the ideas under discussion in connection with the experimental networks seemed to be incompatible, and the prospect for the user seemed bleak. But, following a quite unprecedented degree of cooperation between various network authorities, there has been a convergence between the different ideas, with the result that two basic kinds of service are now seen as possibilities for public networks of the future. Existing experimental networks seem likely, in many cases, to adapt themselves to accommodate one or other of these basic services, which have become known as the datagram service, and the virtual circuit or call service.

An example of the datagram type of network is the EIN (European Informatics Network) (Ref 6), while the virtual call network is typified by the Euronet network (Ref 7) which will use switching centres identical in design with those of EIN, but with additional front end processors to implement a virtual call service. The use of a datagram network as a basis to carry a virtual call service is also apparent in other networks such as Telenet in the USA (Ref 8), and Datapac in Canada. The similarity in architecture between these various networks could provide various modes of interconnection and this may prove to

be very important for international data communications in the future. It is worthwhile looking in more detail at these two kinds of service, since it is likely that users on a worldwide basis will be greatly affected by the outcome of discussions that are now taking place.

The datagram service is basically simpler than that of the virtual call. Packets are accepted from users and handled independently of each other, so that all information for guiding a packet to its destination must be contained in a header accompanying each packet. In contrast, the virtual call service, which also handles packets, has a preliminary call establishment phase whereby information is stored about an anticipated sequence of packets from a subscriber. Thereafter, each packet need have only an abbreviated address, because the users local switch adds whatever is necessary to pass it through the network.

From the users view point, there may seem to be little to choose between these services, but different methods of implementation within each national network may yield quite different performance characteristics. This is a subject that is now being hotly debated and on which further experimental evidence is required.

A further topic of debate is whether the virtual call should be implemented only in the switches, to which users are connected, or whether buffers and other resources should be assigned to every user's virtual call at each of any intermediate switches. The latter scheme implies that the route taken by a call is fixed when it is set up, and means that flow control can be carried out easily, so that congestion is much easier to prevent. On the other hand, there are advantages in dynamic, rather than fixed, allocation of resources at intermediate switches, tó allow easy alternate routing within a call, and to make better use of network capacity, especially when the traffic flow is very irregular. These topics should be the subject of simulation and experiment before vital decisions are taken that could affect users for years ahead.

7 The Use of Satellites

The packet-switching networks being built or planned by various countries will provide services more appropriate to data communications than those presently provided by the telephone network, and will also use transmission resources more effectively - a factor particularly important with long-distance international links. Unfortunately, as the number of nodes in a network increases, considerable queueing delays can arise when several nodes are in tandem.

The introduction of satellite stations into a data communications network could enable direct connections to be set up between the extremities so that, although a satellite channel has a transmission delay of a quarter-second, the consequence for the integrated satellite/terrestrial system may be a faster-response network. Furthermore, the multiple-access capability of the communications satellite would enable any number of links to be established without decreasing the transmission efficiency of the satellite repeater. This would make it possible to use earth stations of modest size, which are cheap to build and easy to locate, to set up a satellite network of considerable capacity.

Much experimental work has been done in the USA (Ref 9) on the use of satellite stations operated in various novel modes, that are well suited for use in packet switching. But it is not clear whether a public network would use such techniques. The policy with satellite channels in telephone usage has been to treat them as equivalent to a terrestrial link, so that a call may be switched through either, without the user being aware of which type of channel he is using. This has constrained the design of satellite systems to accord with the needs of the telephone network, and it remains to be seen how this will influence their use for data communications.

Recently there has been a proposal to associate satellite earth stations with the nodes of EIN to allow them to exchange data via the experimental Orbital Test Satellite due to be launched for the European Space Agency in mid 1977. The earth stations would be small, unattended installations, with a small parabolic refector antenna of not more than three metres diameter, an inexpensive receiving system and a low-power transmitter. Signals would be transmitted at high speed, using time-division multiple-access (TDMA), where by each earth station sends short bursts of signals, which arrive at the satellite one after the other, so that the repeater transmits one signal at a time, with consequent improved efficiency and reduced distortion.

However, it could be many years before traffic satellites assigned exclusively to data, with techniques designed to employ them efficiently, will become available.

Unfortunately, the provision of common data connection services is only a small part of the solution to the problem of easy interaction between users. Here an analogy can be drawn between communication among computers and communication among people. All people have the same mechanism for speech and hearing, but use a variety of different natural languages which prevent them from intercommunicating. The new packet networks can give a universal transport mechanism for data, but the very considerable differences between system languages and data structures used for computers of different manufacture make them largely incompatible.

The inherent incompatibility between computer systems has always been a serious problem for users, but the advent of data networks has made it much more worthwhile to find a solution. The current approach is to define a set of layered protocols which offer standard methods of interfacing at a variety of levels; it then remains for each of the systems to be adapted to communicate using these protocols in addition to, or possibly instead of, their existing methods. As an example, the user-oriented protocols being developed for the European Informatics Network are shown in Figure 1.

Protocols fall broadly into two categories: those which adapt the communications facilities to have a common interface to users, and those which make users' systems mutually compatible. The first type are shown at the bottom of the figure. They are as follows:-

THE LINE PROTOCOL (LP) deals with problems of error detection and retransmission and detects line and switch failures, it is responsible for ensuring packets arrive error free at the level of the TRANSPORT STATION.

THE END TO END PROTOCOL OR TRANSPORT STATION (TS) is essentially a multiplexer and demultiplexer. It takes the packet stream from the LINE PROTOCOL and examines addresses in order to separate the stream into several substreams which are used by other levels of protocol. It provides services such as the ordering of packets in a substream and may also create calls between processes representing higher level protocols. The Transport Station exchanges 'letters' and 'telegrams' with Transport Stations in remote Subscriber Computers, and it may include facilities such as an interrupt whereby an urgent telegram may be superimposed on a call.

THE NETWORK CONTROL MODULE (NCM) has the task of interacting with the communications facilities and with the network control module at a remote site, in order to monitor the behaviour of these facilities. The NCM may also contain a mechanism for signalling alarms to operators at attached computers, and for allowing them to control a communications network.

Once the end to end protocol has been implemented at a Subscriber Computer (SC) attached to a network, ports are available for a connection or liaison with ports in other SCs. The communication between these ports can be regarded as error free, because problems of retransmission, sequencing etc. are taken care of in the lower levels of protocol. It is therefore convenient to regard each pair of ports as connected directly together with just a delay inserted in the path between them. The line protocol, and the transport station therefore greatly simplify the use of a communications network.

The second type of protocol, intended to make computer systems more able to interact together are shown in the centre part of Figure 1. They are briefly described as follows:-

THE BULK TRANSFER FACILITY (BTF) is intended to control the transfer of large quantities of data between one port and another and to handle problems such as the different word length in the machines that are in communication. It also has to cater for recovery after breaks in transmission, so that the bulk transfer operation does not have to start again from the beginning if it is accidentally suspended.

THE LINE ORIENTED PROTOCOL (LOP) is concerned with an interactive exchange of information between a pair of ports. It is responsible for maintaining calls between ports and takes care of problems of reconnection if breaks should occur, ensuring that the call is restored between the correct correspondants.

THE FILE TRANSFER PROTOCOL (FTP) makes use of the BTF to tranfer a structured file of information from one system to another. The definition of a uniform structure for files exchanged between different systems connected by a network makes it possible to map from the particular file structures used by each system onto the standard FTP.

THE REMOTE JOB ENTRY PROTOCOL (RJE) may be based on the standard FTP. It allows jobs to be accompanied by an appropriate 'job control card' to a remote site, and the results of the job can be eventually received back, also in the form of a standard format file.

THE INTERACTIVE TERMINAL PROTOCOL (ITP) is based on the existence of a line oriented protocol. It allows the terminals wishing to communicate through a network to do so with a standard set of commands and responses.

THE VIRTUAL TERMINAL (VT) is able to perform an arbitrary set of terminal-like operations, typical of a range of real terminals that might be connected to the network. Handlers in each of the centres' systems can be written to manipulate the virtual terminal and this allows the interaction between services in Subscriber Computers, and terminals that communicate with them through the subnetwork to be made in a standard way. It is then necessary to associate a terminal handler with each type of physical terminal, in order to translate the real terminals' commands and responses into those defined for the virtual terminal.

A NETWORK CONTROL LANGUAGE (NCL) or network command language is highly desirable to allow users to communicate with the network services in a standard manner. It is, however, difficult to define a universal language, and a number of application oriented languages may be the best practical way to suit different users.

SPECIAL APPLICATIONS PROTOCOLS (SAP) for more advanced forms of interaction may be required later. An example would be a protocol for handling interactive graphics terminals in a uniform manner.

The agreement and implementation of standard high level protocols similar to those described briefly above makes different systems look similar to users, and this approach could be extended to still higher levels of protocol. For example, there has been a gradual development of techniques to allow data to be shared by a number of users, operating under one single information management system, and it is now becoming worthwhile to develop distributed database management systems, so that databases in geographically widespread locations can be used by a wide range of terminals and computer systems. The demand created for network based services, made possible by developments of this kind, may well require considerable changes in the structure of future large computer systems.

9 The Data Network Boundary

For historical reasons, communications facilities are provided by Authorised Carriers (usually the PTTs in Europe), while data terminals and computers are supplied by the computer industry. There is no clear natural boundary in a computer network where communications give way to computing and this gives rise to conflict which distorts the evolution of teleprocessing systems.

At first, Carriers generally regarded data as a nuisance, but now are increasingly seeing the provision of a public data network as a vital extension of their business, and are taking a major role in the definition of standards. The impact of this on future developments is not yet clear, but an indication is given by what has happened with the European Informatics Network and the Euronet Project.

When the European Informatics Network was defined, the Carriers' discussions on packet switching were mainly concerned with a datagram service. EIN was, therefore, designed to give such a service, but had some additional features anticipating a possible virtual circuit service. However, the main features of a virtual circuit are provided by the EIN Transport Station protocol described above. This protocol is, of course, implemented by software in users' systems.

By the time Euronet was planned, the virtual circuit definition had been improved, and the European PTTs (who are providing for Euronet a communications network to link terminals to various data bases) adopted the EIN design, but added terminal processors offering the CCITT X 25 virtual circuit interface (Ref 10). The first type of protocol defined for EIN and described above, is not, therefore, required by Euronet users, because X 25 offers a similar scheme in the network rather than in the users' systems. Thus, the boundary between the communications facilities and the computing services has moved towards the user, during the interval between the design of EIN and the design of Euronet. This is illustrated by figure 2, which should be compared with figure 1. As the detailed design of Euronet continues, it seems likely that some higher level protocols such as the Virtual Terminal may also be implemented within the terminal processor. For example, to handle character terminals a Packet Assembly Disassembly (PAD) function has been proposed to interface them to the packet network. This PAD can readily be augmented to include simple VT features suitable for a range of terminals.

The disadvantage of this trend is that the opportunity for the user to develop novel systems may be reduced if the communications functions become stereotyped. It is therefore important to provide a means to by-pass the higher level facilities offered by a network, and the use of the basic datagram service is one method by which this may be done.

10 Future Prospects

The move towards including, within a data network, facilities that initially formed part of the users computer system seems likely to continue. Until recently, the development of big computer systems was determined by the large main frames, and the software that their manufacturers built around them. The early facilities for data communications were therefore tailored to the needs of such systems, within the constraints placed by the use of the telephone network. But the advent of micro processors is allowing distributed logic to be used increasingly, and the agreement on standard protocols permits many basic communications functions to be realised more cheaply as common facilities in a public data network, rather than as part of the users system. This will leave the users to concentrate on high level functions that are special to the services they use or provide.

The advent of techniques such as Viewdata, means that the era of the home terminal may not be far ahead (Ref 11), and it is difficult to see how this can be provided and maintained other than by the traditional carriers. Hopefully this will lead to intense competition within the computer industry to provide services that suit the needs of the small user, who should form a mass market more valuable than that of the present large organisations that use today's data services. This will truly cause a revolution in the way that future teleprocessing systems must develop.

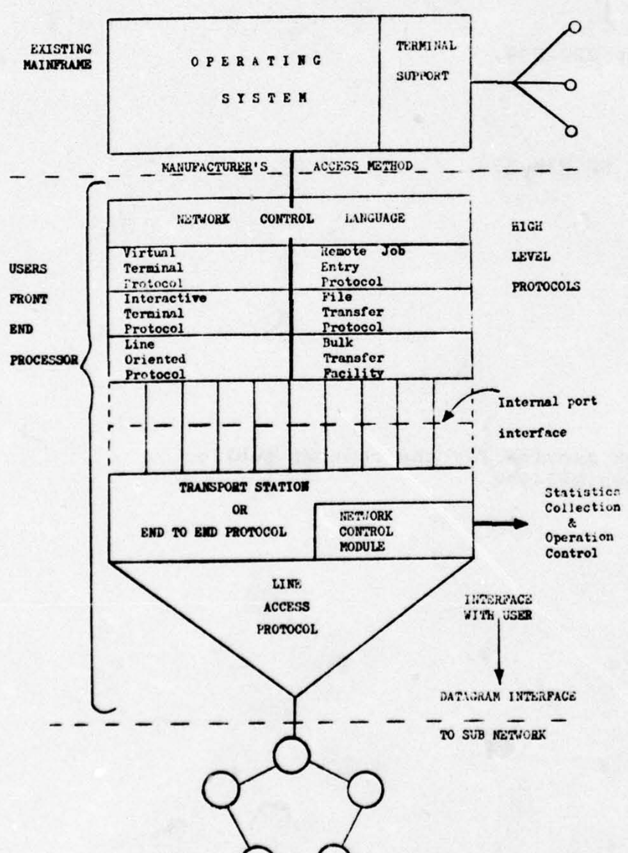


FIG 1 POSSIBLE EIN PROTOCOLS

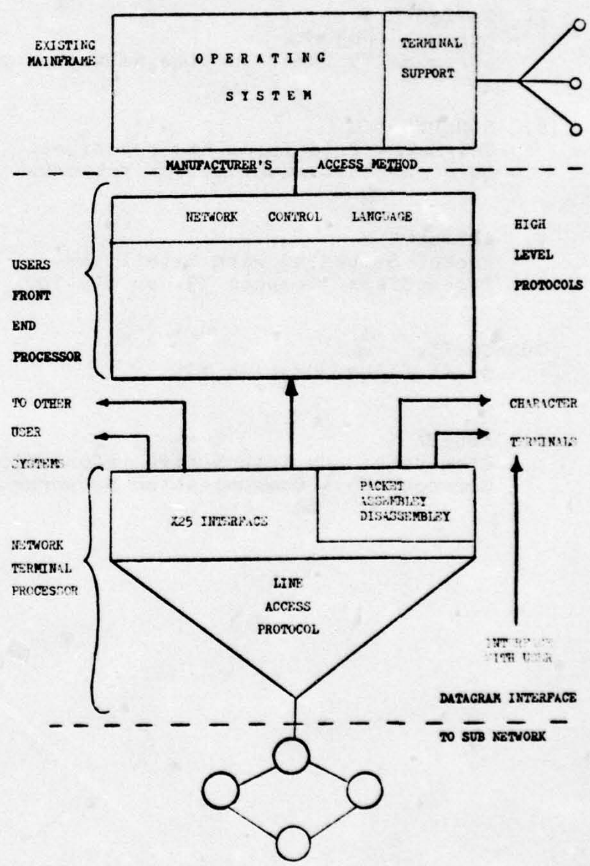


FIG 2 EURONET SWITCHING CENTRE

REFERENCES

1. ARNDT G & STEMBERGER K
   Practical experience with the telephone used as a data terminal.
   Eurocomp 75 - Communications Networks. pp 161-175

2. HAUG T
   Data Communication in Sweden.
   Proceedings IIASA Conference on Computer Communications Networks, October 1974. pp 51-67

3. CCITT
   4th Plenary (white book) Vol VIII 1969.
   Recommendations XI, X2, X20, X21, X50

4. ROBERTS L G & WESLER B D
   Computer Network Development to Achieve Resource Sharing.
   Proceedings AFIPS SJCC 1970. pp 543-549.

5. DAVIES D W
   Principles of a data Communications Network for Computers and Remote Peripherals.
   Proceedings IFIP Congress Edinburgh 1968. pp709-714.

6. BARBER D L A
   The European Informatics Network: achievements and prospects.
   Proceedings ICCC76. Toronto, October 1976.

7. DAVIES G W P
   Euronet Project.
   Eurocomp 75 Communication Networks. pp 229-239.

8. ROBERTS L G
   Telenet - Principles and Practice.
   Eurocomp 75 Communications Networks. pp 315-329.

9. ABRAMSON N
   Packet Switching with Satellites
   Proceedings Compcon 73. pp 695-702

10. CCITT.
    Draft Recommendation X25

11. FEDIDA S
    View data: an interactive information service for the general public.
    Eurocomp 75 - Communication Networks.pp 261-282

Maximising the Use of an Information Service in an International Environment.

W.A. Martin

European Space Agency

Space Documentation Service

ESRIN, Via Galileo Galilei

00044 Frascati, Rome, Italy

SUMMARY

The development of the ESA Space Documentation Service from 1964 to date is briefly reviewed. SDS database policy, which must satisfy the needs of both the Agency and its member countries, is explained and utilisation trends for all major databases examined. Based on a target of self-support since 1971, the evolution of SDS charging policy and the integraion of the recently introduced RTC (remote terminal concentrator) are described. The derivation of database related costs, their potential for reduction, and the useful price reductions which would result from a significant increase in overall system load factor are outlined. The consequences and implications of working in an international environment are reviewed and the particular problems of data communications in Europe are emphasised. The exponential growth in demand for information services as indicated by recent projections is noted, highlighting the need to improve and simplify current types of service; ongoing experimental work at SDS on an integrated information base are briefly mentioned including some thoughts on the multi-lingual requirement.

1. INTRODUCTION: SDS AND ESA RECON

The European Space Research Organisation (ESRO)* was established in June 1962, following the work of the European Preparatory Commission for Space Research (COPERS) on signature of a convention which entered into force in March 1964. During the COPERS period a group of experts had recommended the establishment of a comprehensive European mechanised aerospace information service and the negotiation of an information exchange agreement with NASA. The Space Documentation Service (SDS) was set up in 1964 and an agreement was negotiated with NASA in the same year.

Some milestones in the development of SDS have been:

1965 - pilot operation of computer retrieval on the NASA tapes under contract;
- input processing: section is created for acquisition, abstracting, indexing, and microfiche production/reproduction (the latter under external contract) of European aerospace literature.

1969 - the NASA File has grown to over 400,000 bibliographic items and off-line searches take 20 hours of computer residence time to run a batch of 30 questions, the operation is now prohibitively expensive;
- following the development by Lockheed of the NASA RECON software under a NASA contract, and an evaluation of available packages, ESRO also concludes a contract with Lockheed (with NASA approval) for the provision of the ESRO RECON software;
- ESRO RECON is implemented on the Organisation's largest computer, an IBM 360/65, at the satellite tracking and control centre (ESOC), Darmstadt, with remote terminals at Head Office, Paris and ESTEC in the Netherlands.

1970 - first Member State national documentation centre instals direct on-line access to ESRO RECON from the Technology Reports Centre, St. Mary Cray, U.K.;
- first demonstration of long-distance remote access via RECON is sponsored by US AEC using an ESRO RECON terminal in Paris to interrogate a computer at Lockheed's Palo Alto laboratory by means of the then new TAC-5 transatlantic telephone cable.

1972 - USAEC RECON terminal at Washington ASIS conference interrogates the ESRO RECON database at Darmstadt via the INTELSAT communications satellite.

1973 - central database is augmented to include COMPENDEX, METADEX, GRA and Nuclear Science Abstracts, and exceeds one million bibliographic items on-line;
- SDS re-locates to ESRIN, Frascati near Rome, and IBM 360/50 computer dedicated to RECON and SDS tasks installed.

* *Member Countries: Belgium, Denmark, France, Germany, Italy, Netherlands, Spain, Sweden, Switzerland, United Kingdom*

1974 - the first non-aerospace database - CA Condensates is added.

1975 - European Space Agency is created;
- for the first time a leased-line terminal is installed in the national centre of a non-Member State at CND, Rabat, Morocco;
- 20 remote high-speed terminals now installed in the Member States and both TYMNET and CYCLADES (France) networks now interconnected with ESANET providing access for the first time to inexpensive, low-speed dial-up terminals; installation of the first SDS RTC (remote terminal concentrator) begins, this "first" again being scored by the Technology Reports Centre, St. Mary Cray, U.K. INSPEC database is added;
- CPU is upgraded to IBM 360/65.

1976 - RTC at St. Mary Cray operational and the installation of a remote medium-speed line printer is agreed, followed by other RTC installations at KTH, Stockholm; Royal Library, Brussels; DTB, Copenhagen; whilst discussions leading to the interconnection of the CNUCE (Italy) and CTNE (Spain) networks reach an advanced stage; exchange of letters between the Commission of the European Communities and ESA for practical cooperation in establishing EURONET is negotiated. ISI Science Citation Index and PASCAL (Bulletin Signaletique) databases added;
- 16 IBM 2314 disks (critical to response time) replaced by 6 high-speed high density disks;
- satellite communication now permits access virtually anywhere in the world, preparations made for ESA RECON demonstration in Bombay at the request of UNESCO.

2. THE SDS CENTRAL DATABASE

2.1 SDS database policy has to satisfy two quite separate needs:

- First to ensure the availability on-line of those databases needed by the Agency* in the support of its various projects.

- Second, to satisfy the aerospace and aerospace-related information needs of the ESA Member States.

2.2 The provision of computer searches of the NASA File to the Member States of ESRO and ELDO was originally the very "raison d'etre" of SDS. Thus this database, to which SDS contributes input in machine readable form. was the first and foremost computer readable SDS resource, and it remains so today.

2.3 However, because no multi-disciplinary mission-oriented information system can cover one category within its scope to the same depth as a discipline-oriented service covering that category exclusively, it is usually necessary to augment the primary mission-oriented database with a number of discipline-oriented databases, each covering a part of the main area of interest.

For this reason SDS augmented the NASA File, as soon as capacity was available, with the databases listed at Table 1 section A. Other available machine readable databases are continually under consideration, in the light of the changing requirements of the Agency. Most recently this procedure has resulted in the selection of the databases listed at Table 1 section B, whilst Spacelab requirements can be expected to give rise to further additions. The verification of these decisions is given at Figure 1 which shows the utilisation trends in actual connect hours for those databases accounting for over 85 per cent of Member States use of ESA RECON. It is significant that NASA recently adopted a similar policy following an investigation of internal user needs, and plans to add new databases during 1976.

It should be noted that not all SDS databases are necessarily accessible to all users in all Member States. Access is defined for each individual password.

2.4 Partly as a result of the introduction of the policy of cost recovery for SDS and partly due to the overwhelming demand for the Chemical Abstracts database, the concept of non-aerospace databases was introduced calling for "sponsorship". The sponsor, who could be the database supplier or an interested user group was required either to pay to ESA the full cost of file creation and subsequent on-line storage and updating, or to guarantee to pay the difference should utilisation be insufficient to recover costs. In the case of Chemical Abstracts the former arrangement was agreed.

* *By "the Agency" is meant the Agency's staff at its Head Office in Paris; and its establishments at ESTEC, Noordwijk; and ESOC, Darmstadt; also its prime contractors, their co-contractors, and scientific groups associated with ESA scientific missions.*

Table 1

A. Databases chosen initially to augment the NASA File

COMPENDEX	Engineering Index
NTIS	US Government Report Announcement
METADEX	Metals Abstracts
INSPEC	Physics, Electrical & Electronic Engineering, Computers and Control

B. Databases chosen more recently to augment the NASA File

WAA	World Aluminum Abstracts
ESI	Environmental Science Index
ISI	Science Citation Index (Physics, Chemistry & Engineering sections)
PASCAL	Bulletin Signaletique (Aerospace-related sections)

C. Non-Aerospace Databases

CA CONDENSATES	Chemical Abstracts Service

2.5 The physical, chemical and engineering sections of the ISI Science Citation Index were brought on-line during May 1976. Over 560,000 references, from June 1972 were made available. This file will be updated each month under the terms of a 15 month experiment during which the extent to which SDS users augment traditional searches by citation, and search ISI by title-word or author, will be evaluated.

2.6 The aerospace and related sections of the French PASCAL database (Bulletin Signaletique) were scheduled to be on-line by mid-1976. Interest in this database stems, not only from the widespread French access to SDS now possible via the CYCLADES network and from Belgium and Switzerland; but general interest has been expressed in this database by many Member States as a result of its useful coverage of the literature of many East-European countries, the Soviet Union, and some Arab countries. SDS is particularly interested in the promised tri-lingual version of PASCAL, in French, German, and English.

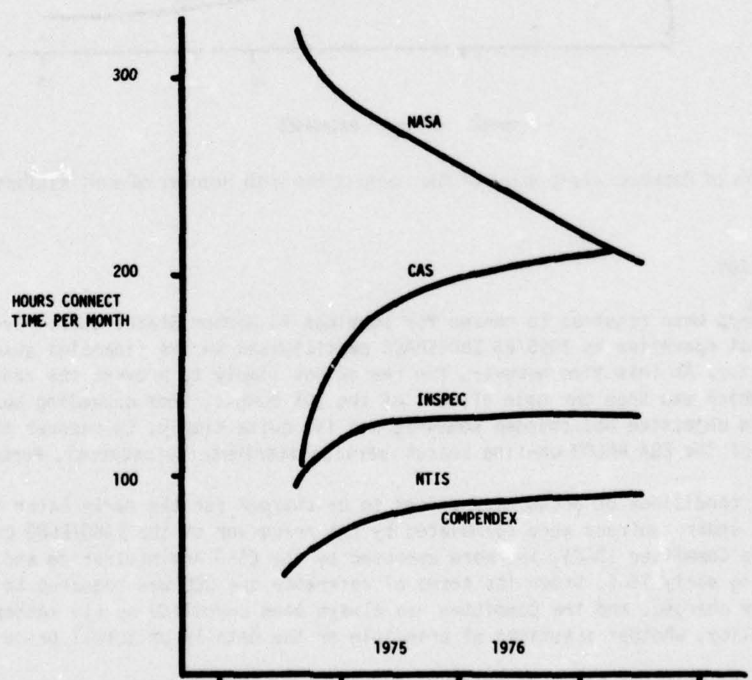


Fig.1 ESA RECON Member States utilisation trends 1975-76

2.7 Accurate determination of the off-line capacity available for file updating is a rather complex matter in terms of the finite number of databases whose retro-file sizes, update file sizes, record sizes, and record complexities bedevil any attempt to achieve a standard approach. SDS has begun to consider such projections, however, since a knowledge of the optimum database content is becoming necessary to proper planning decisions. The "unit database" is taken to be about a half million items on-line updating at around 5,000 items per month. First order estimates indicate capacity equivalent to some 25 to 30 unit databases, and, allowing for a degree of miscalculation and the necessary contingencies, would indicate an available capacity for some 8 to 10 additional unit databases.

Figure 2 shows the influence on unit costs of file maintenance as more databases are added within available capacity. Optimisation of this capacity will result in a lowering of existing prices to users and should also contribute to increasing system load factor thus increasing the centre's turnover.

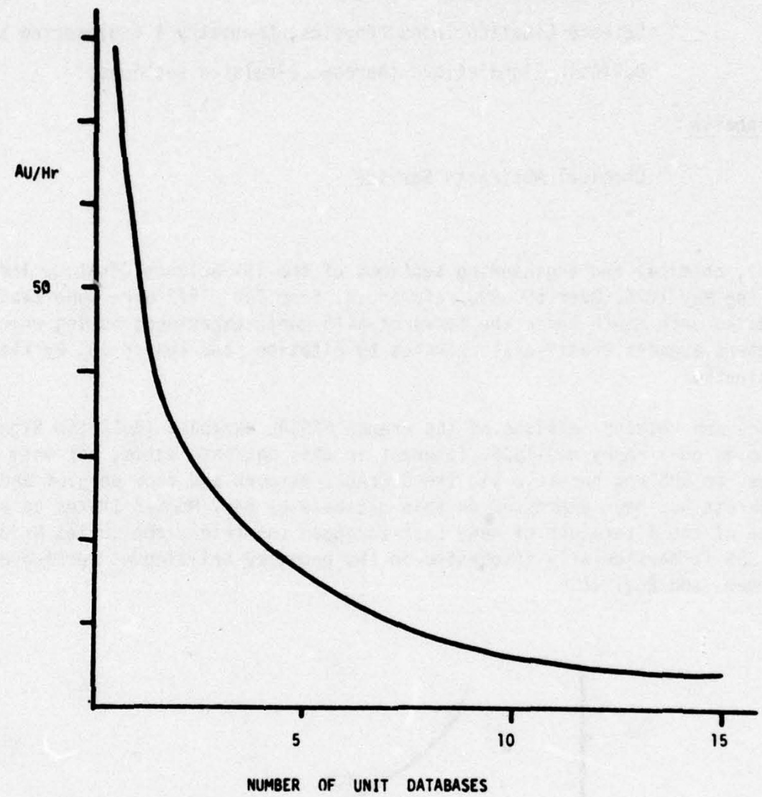


Fig.2 Variation of database component of file connect fee with number of unit databases maintained

3. SDS CHARGING POLICY

3.1 SDS has always been required to charge for services to Member States users. Even for the first year of pilot operation in 1965/66 EUROSPACE participated in the financial guarantee required by the contractor. At this time however, the reason was simply to prevent the annual costs of the contract, which was then the main element of the SDS budget, from exceeding budgetary provisions. Nowadays the objective has changed somewhat and is, quite simply, to recover that part of the total cost of the ESA RECON on-line search service attribute to external, Member States use.

3.2 The initial conditions of access and prices to be charged for the early batch search services carried out under contract were formulated by the precursor of the ESRO/ELDO Documentation Consultative Committee (DCC), and were approved by the ESRO Administrative and Finance Committee (AFC), during early 1965. Under its terms of reference the DCC was required to advise on the prices to be charged, and the Committee has always been consulted on all matters relating to charging policy, whether questions of principle or the details of actual price levels.

3.3 It was in 1970 that the Director General's policy statement mentioned for the first time the possibility of placing SDS services to Member States on a self-financing basis and modest success in earning income during that year led, in 1971, to the introduction of the policy of cost-recovery. However, by early 1972 the D.G., said that he thought it necessary to recognise that if the

service were restricted to the aerospace community it would never be properly economic.

3.4 Although the first external terminal, at TRC near London, had been in operation since the beginning of 1971, and two others at CNES Bretigny, and ZLDI Munich, were installed during that year, SDS notions of charging policy were as yet rudimentary and the initial financial arrangements were of necessity somewhat ad hoc.

By the beginning of 1972, however, a working group set up by the DCC had begun to examine the problem and had already noted that the practice of Member States paying the full cost of the leased telephone line from ESOC to the national documentation centre was inequitable since those countries located, by geographic accident, furthest away were penalised. Based on the limited experience then available it seemed that external users with the possibility of serving a reasonably large clientele (e.g., national or regional documentation centres, and major industrial groups) were prepared to pay about 20,000 AU as an annual fee to ESRO in addition to local line and equipment costs. On this tentative basis the cost recovery gap (then estimated at about 400,000 AU pa)* could theoretically be closed if the SDS network could include about 20 external leased-line terminals (LLT) in addition to about 5 ESRO terminals; this was thought to be technically feasible.

3.5 At its meeting in March 1972 the DCC took its first decision relating to the charging policy for direct on-line access to the ESRO RECON service, and approved charges of: 11,000 AU pa - rent and maintenance of terminal (and allocation of central resources), and 22,500 AU pa - ESRO SDS connection fee.

For this total of 33,500 AU (or about $ 40,000) SDS was prepared to guarantee a minimum of 1200 hours RECON availability, i.e., a unit cost of about $ 33 per hour. These figures assumed that ESRO would provide a tele-communications network reaching into each Member State requesting the RECON service, whilst the terminal was a sophisticated device utilising a video display at 240 char. per sec., and incorporating a low-speed printer. The central database then comprised four files totalling almost one million bibliographic items.

3.6 The identification of potential customers in Europe able, and willing, to spend lump sums of such magnitude for an information service, however advanced, was an uphill task, and by the following year SDS had produced proposals for the sharing of the connection fee over two or more terminals, in an attempt to lower the entry fee to RECON. Though this option did enable a group of industrial companies in the Netherlands to gain direct access to SDS, in general these cost-sharing schemes were unsuccessful; what was needed was a "pay-as-you-go" system.

3.7 Following the move of SDS to ESRIN, Frascati in 1973, the installation of an IBM 360/50 computer dedicated to RECON support and associated tasks, and anticipating the introduction of new computer statistics and accounting routines during 1974, the concept of hourly file connect fees had been explored by SDS and a tentative scale of prices derived ranging from 28 to 38 AU ($ 34 to 46) which could be applicable to both LLT and TTY terminals. These prices were approved by the DCC in February, 1974, and were increased slightly at the end of 1975 (see appendix 3).

3.8 In deriving these early file connect fees SDS did not then have detailed statistics of utilisation nor breakdowns of total operating expenditure in the form needed. It is remarkable, and probably coincidental, that in the subsequent refining of the early costs, they were not changed significantly. The procedure followed was first to calculate the actual cost to SDS of maintaining a given file, including lease fees, programmer's time, computer time, mass storage, etc., then to estimate the probable proportion of total on-line connect time spent on the file; hence the hourly rate. No comment is needed on the primaeval forecasting techniques then available, whilst the derivation of costs left much to desired.

3.9 Forecasting is now based on detailed statistics of file utilisation (see appendices 1, 2 and 4) which enable us to plot accurate trends and to project these with considerably more confidence (Fig. 1). Paradoxically we have moved away from the early attempts at detailed cost accounting to derive database maintenance costs, and now favour a simple, top-down approach. Since, if a file is kept on-line but never used, none of the network resources and virtually none of the central computing on-line resources are used, the costs actually incurred will be those associated with mass storage and with periodic off-line file updating. These database related costs are identified as part of the total annual operating costs and are allocated across the various databases on the basis of total file size, update file size, average record size, record complexity, and the number of inverted file entry points. Since we may in future introduce additional processing as part of file updating this will also have to be accounted for. Current SDS prices are shown at appendix 3, and an example of a computer-prepared monthly invoice for an LLT user is shown at appendix 5. Note that the amount in AU is converted to the member

* *Whilst operating RECON on the ESOC computer as a moderate priority task.*

country's national currency of payment. A separate statement of royalties incurred is prepared (appendix 7) showing an amount in US dollars (for all databases to date).

3.10 Up to now the figure of 35,000 to 40,000 connect hours annually has been the basis assumed for cost-recovery and all file prices are calculated on this basis. By the end of 1975 installed network capacity had reached about 35,000 hours, 10,000 of which was used by the Member States, and 6,000 by ESA and SDS. By the end of 1976 network capacity will comfortably exceed 100,000 hours. Assuming that the load factor does not change significantly some 30,000 hours of Member States use could result. We believe, however, that the capacity of the SDS central computer is greater than 100,000 hours; if we could push utilisation up towards this figure there is a tremendous scope for really significant price reductions. An important factor in achieving this goal will be the addition of more databases to increase the load factor.

3.11 As the concept of networking and inter-networking gains ground and is realised operationally we shall see a move away from a small number of centralised on-line information centres to a proliferation of, perhaps smaller or more specialised, centres. It is to be hoped that, in the complicated planning of such information complexes the optimisation of the individual centres will not be overlooked in the attempt to optimise, at a higher level, the overall system represented by the networks and interconnected hosts. If they are overlooked it is conceivable that any economies which might have been realised as a result of physical networking will be offset by the prices charged by the hosts in their attempts to achieve financial viability.

As an extreme example of what could happen, current SDS prices, which are expected to result in operating cost-recovery quite soon, range from 32 AU to 56 AU ($ 42 to 73) per connect hour, the lowest price being for the NASA File. If, as has been suggested quite seriously, SDS were to remove all databases other than the NASA File we calculate that it would be necessary to charge 135 AU per hour ($ 176) to recover costs. In fact, of course, such a price would be counter-productive and in practice it could prove quite impossible to become financially viable. *

Conversely, if more databases are added unit costs of file maintenance will reduce whilst system load factor can be expected to increase. Under these conditions it is conceivable that existing prices could be halved!

3.12 The introduction of the RTC (remote terminal concentrator) by SDS recently, permitting access by low-speed TTY data terminals, has given rise to new aspects of charging policy. The arrangements approved by the DCC were that the national centre installing an RTC shares the hardware and installation costs equally with SDS, (i.e., a non-recurring cost of 10,000 AU); and provides local telephone lines and low-speed modems. The centre issues contracts to its users, handles invoices (prepared centrally by SDS, see appendix 6), and is responsible for promotional and training activities (though the latter are supported by SDS, to the extent possible, particularly during the first year).

4. THE INTERNATIONAL REQUIREMENTS

4.1 ESA currently has ten full Member States, and four Observer States ** two of which are participating in a special ESA programme. In addition SDS also has contact with other non-Member States. In all some nine different European languages (not to mention cultures) are involved! This gives rise to the need for multi-lingual subject specialists to abstract and index documents destined for announcement in NASA STAR. We ensure, however, that these people do not spend more than a proportion of their time on input processing tasks, leaving them free also to carry out user liaison functions. These activities are diverse and include a particularly important ESA RECON advisory and training role, first when a Member State is considering which new services it could introduce nationally based on SDS, and secondly in support of promotional and training seminars in connection with dial-up access. Language barriers, though potential obstacle have never in fact presented a serious difficulty to SDS. With the establishment of SDS "windows", i.e., national centres, in Member States the problem tends to recede even further.

4.2 The GNP and population of ESA Member States vary significantly, resulting in widely varying requirements in the level of services needed. This spectrum of need has led to the early installation, followed by multiplication, of leased-line terminals giving a very high-level service in some countries, whilst others until very recently could not justify on-line access at all. However, the availability of dial-up access at very low cost now enables all countries to gain access.

* *see 3.3*

** *Austria, Canada, Ireland, and Norway. The Republic of Ireland signed the ESA convention in December, 1975 and will become the eleventh member of the Agency once the new Convention is ratified by all existing member states.*

An extremely important area of responsibility for the DCC has been to advise on the projection of SDS services throughout the Member States. Almost inevitably this led a number of Members to become instrumental in introducing into their countries direct on-line access to the SDS computer.

It will be realised that as SDS services developed and extended, the tasks of the individual DCC members became progressively more onerous, as they were called upon to reconcile the sometimes conflicting policy and financial requirements of (ESRO) ESA with those of their own national services. More recently the situation has been complicated further by the emergence of the EURONET concept within the Commission of the European Communities, which has called for the harmonisation of national information policies in the CEC member countries with the international plans of the CEC itself, and with those of ESA whose member countries are not wholly coincident with those of the CEC! Small wonder, then, that full agreement cannot be reached overnight.

4.3 It has been the experience of SDS that the DCC has functioned quite exceptionally well, never lacking the necessary goodwill and understanding to achieve joint agreement and has been a very positive example of how international cooperation can result in the realisation of a common, cost-effective resource. Following the creation of ESA, and recognising that the extent of SDS service throughout Europe must have a significant impact on the European information environment, the DCC was superseded in December, 1975, by the ESA Documentation Advisory Group (DAG), a delegate body which met for the first time in June of this year.

4.4 The first network interconnection undertaken by SDS was the unidirectional link with TYMSHARE which enabled TYMNET users to access the SDS databases at Frascati and provided the first experience of dial-up access to ESA RECON. It is interesting to note the policy adopted by one member country, Switzerland, which has opted to use the Lausanne TYMNET node as the Swiss national access point, monitoring traffic density to provide data on which to base any future decision.

In contrast the decision of France was to connect the French national CYCLADES network with SDS and this network now provides the approved means of access to SDS for French users, though some established TYMNET users in France insist on maintaining their use of this latter route at present. Similar interconnection is now under development with the CNUCE Italian Universities Network, whilst possible interconnections with the Spanish CTNE and Nordic SCANNET networks are under discussion.

A dialogue has long been in existence between the Commission of the European Communities and ESA and has resulted recently in the drafting of letters of agreement relating to scientific and technical information policy in the two organisations. SDS is scheduled to become one of the first two hosts to be connected to EURONET, the other being the Deutsches Institut für medizinische Dokumentation und Information (DIMDI) at Cologne. It has been established that the minimum service to be provided through EURONET from ESA will be based on those databases on-line from SDS at the date of the first operational service through EURONET.

4.5 Data communications pose special problems in Europe, not only because of the very high cost of line rentals as compared with, say, the United States, and the heavy surcharges imposed should the user be able to improve the efficiency of line utilisation by operating at higher speeds, but also because of the problems involved in attempting to co-ordinate the efforts of so many (currently 11 for SDS) autonomous P.T.T.'s, each of which, in spite of CEPT, offers services of differing quality, interprets common standards in the most uncommon manner, and communicates in a different language, no one of which, it appears at times, is an adequate basis for communication between any two P.T.T.'s!

The international high-speed trunks between Frascati, Italy, and Darmstadt, Germany, are absolutely vital to RECON operation, carrying the bulk of network traffic. Overall availability of these trunks is however, no better than 85 per cent! This is about the best than can expected from international leased data-circuits in Europe at present. When central computer and peripheral down-time is superimposed on data-circuit outages the net ESA RECON system availability seen at the user's terminal is no better than 80 per cent. Thus fact has prompted the commissioning recently of a study of the reliability of the entire ESA RECON system. The study is intended to highlight the areas where improvements must be made to realise the required level of 95 per cent reliability at the remote terminals, and should identify not only the faulty or unreliable components of the system, but also the reasons for the failures.

5. ADVANCES IN RETRIEVAL TECHNOLOGY

5.1 It is interesting to consider the recent trends in information use as indicated by the growth of on-line searching. Figure 3 shows the trend from 1971 to 1974 in the United States (figures due to Prof. Martha Williams) and in Western Europe (EUSIDIC). The U.S. curve would appear to be exponential. Figure 4 includes two projections of use in Europe, one obtained by superimposition

of the points of origin of the Martha Williams curve over that for EUSIDIC, the second was prepared by PA Management Consultants under a contract with the European Commission[1]. They may be considered to be the result of optimistic/pessimistic assumptions but whichever is nearer the truth, explosive growth in utilisation is indicated.

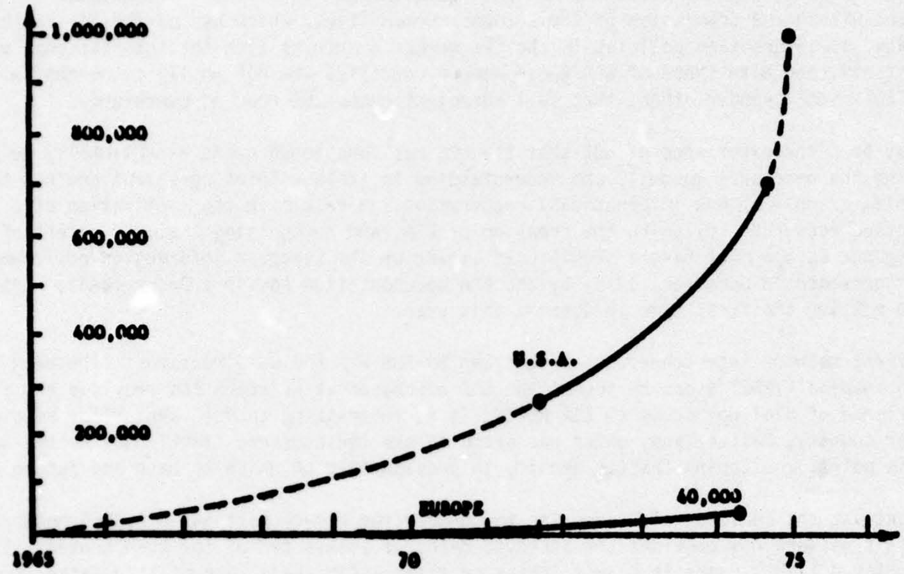


Fig.3 Growth of interactive on-line bibliographic searches in the USA and Europe: trends (USA Professor Williams; Europe EUSIDIC)

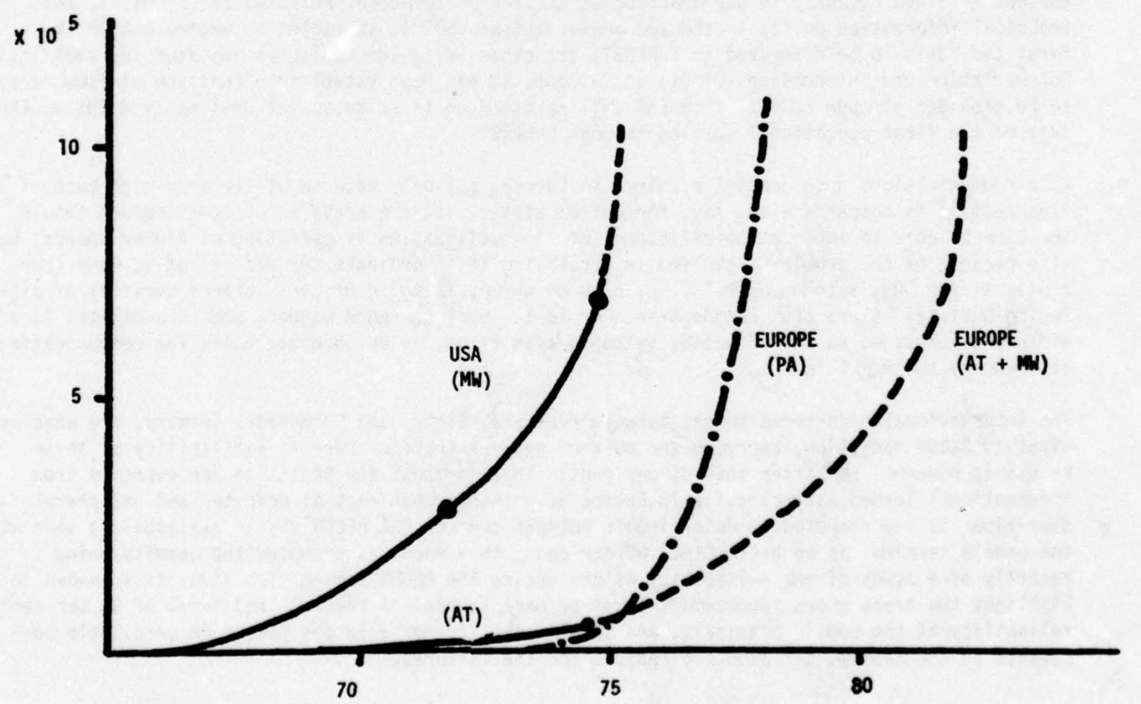


Fig.4 Growth of interactive on-line bibliographic searches in the USA and Europe: *trends and projections*

5.2 It has been said that "the future of the knowledge industry and all its components, without exception, is closely related to the automation of information. The unanimous view (of Delphi surveys) is that during the decade 1980 - 1990 automated information will entirely replace the more or less adequate manual processes at present transmitting and disseminating knowledge". Further: "The evidence shows that the experts have not even begun to give any serious consideration to more complex applications, or to types of services which may become quite common in the future. It would seem that barely ten years after the first automated information systems were introduced, we are already prisoners of a fixed line of thought."[2]

5.3 Can we meet this challenge? In common with most other information centres SDS currently treats each database as a discrete information source. The user, who may need to search several databases in order to obtain an exhaustive answer, must develop a different search statement for each database, making allowances for the differing vocabularies, indexing philosophies, or possibly a lack of indexing. The process is time-consuming, inefficient, and frustrating to the searcher who runs the risk of failing to select the one database which would be most productive.[3] It is also entirely artificial since there are simply not that many different kinds of information (over 400 computer readable databases have been reported recently[4]). The artificial nature of this approach is well illustrated by the degree of overlap which exists between many secondary services.

5.4 Contractual arrangements with the database suppliers, who tend to charge royalties on the basis of file connect time and are also understandably concerned to see that the often unique features of their database should not be lost or diluted in an on-line system, are partly responsible for the current unsatisfactory state of affairs. The information centres themselves must also take their share of responsibility, however, since none has yet produced a solution.

5.5 The need for the development of improved techniques for the simultaneous searching of many bibliographic databases has long been recognised by SDS. However, earlier studies have served only to emphasise the difficulties rather than to suggest possible solutions. All approaches tended to be based on the notion of cross-linking thesauri, a notion which is operationally unattractive.

5.6 During 1975 SDS initiated[5] an experimental project as part of its investigation of the integral information base concept.[5] This concept is based on the following requirements:

- each source database as received from the supplier will be maintained as a discrete datafile on the computer, retaining fully its identity and any unique characteristics

- each source database may be searched individually if the searcher so wishes

- the user will see, through his remote terminal window, as default option, an integral information pool (irrespective of source databases) which may be searched by the development of a single search statement on a common inverted file

- the user may also select one or more discrete databases prior to searching, i.e., any subset of the total information base.

5.7 Software, both off-line and on-line, already exists to support the integral information base approach, but to be realised in practice a common retrieval vocabulary for all merged files is necessary. In addition to this common vocabulary each database would be augmented with any special retrieval keys which may be provided by the database supplier. A developed form[5] of free-term indexing could provide this common retrieval vocabulary.

5.8 This developed form of free-term indexing presupposes improvements in three important areas. First, that the inverted file would be derived using a machine-aided approach in order to eliminate debris, correct misspellings, amend spelling variants, standardise abbreviations, and minimise the singular/plural problem. Second, that term phrases would be introduced, derived again via a machine-aided approach, and based on some variant of siginificance analysis. Third, that some form of search-aid would be developed, comparable with the Related Terms feature of a controlled vocabulary, called Associated Concepts. The derivation of Associated Concepts could be expected to be on the basis of statistical associations of terms actually existing in a large database.

5.9 As the first phase of this experimental project SDS wished to develop all the necessary techniques on a single database, and preferably the one with which we were most familiar. NASA obliged by providing two years of NASA STAR abstracts tapes. The experimental database is already on-line, and the initial inverted file of title and abstract terms has been completely intellectually processed and fed back to form a Reference Term File which will be the basis of the on-line inverted file. A technique based on the notion of extentional relationships developed by Wessel[6] is being used to derive a variety of possible formats for the Associated Concepts (RT analogue)

search-aid. We may not be able to incorporate term phrases in this first phase. A group of experienced ESA RECON users are being asked to evaluate the resulting capability, under the direction of an external co-ordinator who will also analyse the results of test searches and produce a final report. We shall thus obtain an objective assessment of the work to date from an expert in the field.

If the results indicate that this line of development is worth following we then hope to repeat the exercise on a second database when it will be immediately possible to test the full range of integral information base features. The results could then lay the basis for a new operational service from SDS.

5.10 The conversion of an experimental success into a useful operational service would probably be dependant on two main factors. One would be the willingness of the database suppliers to jointly seek alternative but equitable bases for the recovery of royalties or use charges. However, if the technique can be shown to be a significant improvement in service to the ultimate user we do not believe that the database suppliers would wish to block such progress.

The other factor would be the extent to which the integral information base concept would be compatible with the increasing demand for multi-lingual systems. We believe that a feasible approach would be to have, initially, titles available in all system languages and to create independant inverted free-term files together with auxiliary search aids for each system language. It is felt that availability of the search vocabulary alone (whether controlled or not) is insufficient since interaction with the database, so vital to effective on-line searching, would be inhibited if titles were not available in the searcher's own language. This approach would provide an immediate multi-lingual capability. The overheads in additional mass-storage could later be reduced as improved techniques were developed.

5.12 I leave you to judge if all information centres are necessarily "prisoners of a fixed line of thought", and to consider whether the apparent level of effort currently allocated to the development of new techniques in automated information retrieval is going to be sufficient to enable us to satisfy an exponentially increasing demand for an exponentially increasing supply of information in all sectors and at all levels.

REFERENCES

1. P.A. Management Consultants Limited. *Forecast of Users of On-Line Retrieval Services for Scientific and Technical Information in Europe 1976 - 1985. Prepared for the Commission of the European Communities, November 1974.*

2. Anderla, Georges. *Information in 1985.* Paris, Organisation for Economic Co-Operation and Development, 1973.

3. Martin, W.A. A comparative study of terminal techniques. Paper presented to the First European Congress on Documentation Systems and Networks, Luxembourg, 16 - 18 May, 1973, EUR 5058, pp 107-164.

4. ASLIB/EUSIDIC. Databases in Europe. European User Series Number 1.

5. Martin, W.A. Unpublished.

6. Wessel, A.E. Unpublished.

Example of summary monthly statistics of ESA RECON database utilisation

SPACE DOCUMENTATION SERVICE

MAY , 76

TOTAL CONNECT TIME : 1691H 1M

CUSTOMER CONNECT TIME :

FILE	TIME	%
0	0H 2M	0.0
1	202H 52M	25.8
2	200H 26M	25.5
3	34H 16M	4.4
4	77H 38M	9.9
5	1H 20M	0.2
6	79H 7M	10.0
7	14H 33M	1.8
8	126H 40M	16.1
9	5H 54M	0.7
10	5H 9M	0.7
11	10H 8M	1.3
14	0H 1M	0.0
17	20H 16M	2.6
19	0H 17M	0.0
32	8H 34M	1.1
	782H 51M	100.0

Example of monthly summary statistics of overall ESA RECON system utilisation by leased-line and dial-up terminals.

SPACE DOCUMENTATION SERVICE

(NUMBERS IN PERCENT) MAY 1976

LEASED LINE TRAFFIC

PORT	LOGONS	BREAKS	COMMANDS	INVALID	CONNECT TIME
A0 SPAIN	11.2	5.0	7.4	8.5	4.2
40 ROME	3.0	10.6	9.3	2.8	13.6
43 ROME	2.7	10.6	1.6	1.0	9.1
A7 GERMANY-SCANDINAVIA	22.0	22.9	28.8	50.1	26.5
A1 NETHERLANDS	20.8	15.6	15.9	11.0	14.3
A6 GREAT BRITAIN	20.5	9.2	16.1	12.6	9.2
A3 FRANCE	8.8	17.9	7.3	9.1	15.7
A5 MOROCCO	3.5	4.1	5.4	3.3	3.5
A2 BELGIUM	3.7	1.8	2.0	0.9	1.2
44 ROME	3.7	2.3	6.2	0.7	2.7
TOTAL	58.3	81.6	76.2	86.6	82.5

TYMSHARE TRAFFIC

PORT	LOGONS	BREAKS	COMMANDS	INVALID	CONNECT TIME
6A	79.0	100.0	82.3	79.5	82.8
6B	17.7	0.0	15.9	15.9	15.5
6C	3.2	0.0	1.8	4.5	1.7
TOTAL	4.0	1.5	2.1	0.9	1.2

DIAL-UP TRAFFIC

PORT	LOGONS	BREAKS	COMMANDS	INVALID	CONNECT TIME
9A BRUSSELS	0.9	0.0	0.3	0.5	0.3
94 DARMSTADT	51.5	40.0	50.7	37.3	44.5
97 BRUSSELS	0.2	0.0	0.0	0.0	0.0
98 ---.---	0.4	0.0	0.1	0.0	0.0
99 ---.---	0.4	0.0	0.2	0.0	0.1
9B ROME	5.9	6.7	6.7	8.6	6.9
95 DARMSTADT	9.0	13.3	10.1	7.0	9.5
62 DARMSTADT	1.4	0.0	1.9	1.6	2.0
60 ROME	12.3	26.7	14.3	25.9	17.8
64 ROME-1200	9.2	0.0	7.8	8.6	8.6
92 ROME-10	6.7	13.3	6.9	8.1	9.4
6E DATEX	2.2	0.0	1.0	2.2	0.9
TOTAL	17.8	5.6	9.4	3.8	6.5

R.T.C. TRAFFIC

PORT	LOGONS	BREAKS	COMMANDS	INVALID	CONNECT TIME
A7 STOCKHOLM	4.7	0.0	3.1	9.2	3.3
A6 LONDON	60.4	66.7	76.1	80.9	74.0
A4 CYCLADES	35.0	33.3	20.8	9.9	22.8
TOTAL	19.9	11.2	12.4	8.7	9.8

Appendix 3

Current SDS File Fees.

1. **FILES AVAILABLE AND ACCESS FEES**

Access to the various files will be subject to the following hourly fees (in Accounting Units).

FILE	A.U.
NASA	32
CHEMICAL ABSTRACTS CONDENSATES	46
METADEX	38
COMPENDEX	44
ESA - ELECTRONIC COMPONENTS DATABANK	49
NTIS - GRA	38
USAEC - NSA	49
INSPEC	40
WORLD ALUMINUM ABSTRACTS	40
ENVIRONMENTAL SCIENCE INDEX	44

2. **ROYALTIES**

All royalties requested by the file owners shall be at the charge of the Contractor. At present the following hourly royalties have been requested:

FILE	US$
CHEMICAL ABSTRACTS CONDENSATES	4
METADEX	2
COMPENDEX	3
NTIS - GRA	2
INSPEC	10

For the Chemical Abstracts Condensates File a royalty of US$ 0,02 per reference printed "off-line" is also due.

3. **PRINT CHARGE**

A fee of AU 0,05 will be charged for each reference printed "off-line" (except for format 1).

4. **OTHER CHARGES**

i) For the Chemical Abstracts Condensates File an additional fee of AU 10. per hour is due to Stichting Nederlandse Informatie Combinatie (NIC).

ii) For dial-up access via the SDS network exclusively, a connect fee of AU 6. per hour is applicable.

iii) For dial-up access via the Tymshare network, a connect fee of AU 9. per hour is applicable.

iv) For dedicated terminals an equipment rental & maintenance charge of AU 11.000 per contractual year is applicable. The cost of the domestic lines is at the users' charge.

5. **PAYMENTS**

All charges, except the royalties, will be payable in national currency at the official ESA rate.

For 1976 the ESA exchange rate of 1 AU is

Royalties will be invoiced and payable in US Dollars.

Rental of equipment shall be payable in advance; other charges will be invoiced monthly.

Invoices shall be paid within one month of the date of issue.

Appendix 4

Example of monthly ESA RECON utilisation report prepared for each database supplier.

A -- SPACE DOCUMENTATION SERVICE - TOTAL FILE UTILISATION FOR RECON USE

PERIOD: MAY 1976

LE 1 N A S A

REFERENCE NO.: 76/ROY-OW

DATE : 1976/ 6/31

SA, SCIEN. & TECHN. INF. OFFICE
'FICE OF INDUSTRY AFFAIRS
D TECHNOLOGY UTILIZATION
SHINGTON 20546, U.S.A.
'TN: MR A.A. DESIMONE, HEAD DISTR. & EXC.

RATE PER HOUR: US$ 0

1	2	3	4	5
. NO.	U S E R	USED TIME	CUMULATIVES	TOTAL IN $
102		25 h 24	143 h 49	0.0
103		23 h 16	114 h 17	0.0
104		4 h 51	14 h 29	0.0
107		18 h 41	96 h 53	0.0
109		15 h 41	59 h 56	0.0
110		0 h 0	1 h 57	0.0
111		2 h 42	18 h 47	0.0
112		2 h 5	24 h 53	0.0
114		3 h 28	47 h 33	0.0
116		31 h 32	145 h 18	0.0
207		0 h 2	0 h 8	0.0
208		0 h 0	0 h 15	0.0
	*** CONTINUE NEXT PAGE ***	213 h 21	1215 h 12	0.0

Appendix 5

Example of computer generated monthly invoice prepared for ESA RECON leased-line terminal user.

ESA -- SPACE DOCUMENTATION SERVICE - TOTAL FILE ACCESS FOR RECON USE
PERIOD: 3 MAY 1976 - 26 MAY 1976
REFERENCE NO.: 76/ACC
DATE : 1976/ 6/31

[illegible]

VIA DISPLAY

1		2	3	4	5	6
FILES NO.	DESCRIPTION	USED TIME	CUMULATIVES	OTHERS	FEE PER HOUR IN A.U.	TOTAL IN A.U.
1	N A S A	13 h 26	165 h 28		32	429.87
2	CHEMICAL ABSTR.	3 h 4	16 h 15		46	141.07
3	METADEX	0 h 58	7 h 14		38	36.73
4	COMPENDEX	1 h 30	17 h 47		44	66.00
5	ELECTRONIC CMP.	0 h 0	3 h 24		49	0.0
6	NTIS - GRA	7 h 15	69 h 44		38	275.50
7	USAE - NSA	0 h 10	3 h 26		49	8.17
8	INSPEC	3 h 52	35 h 48		40	154.67
9	W. A. A.	0 h 35	1 h 45		40	23.33
11	ENVIR. SC. IND.	0 h 14	1 h 29		44	10.27
		31 h 4	322 h 20		TOTAL :	1145.60 (A.U.)
						671.46 (LSG)

EXTRA FEE FOR FILE 2 10 AU/HR 30.67 (A.U.)

N O T E S :

1. CONTRACT NO. : 13/73
2. DURATION OF CONTRACT : FROM 75-11-15 TO 76-11-14
3. HOURS FORESEEN IN CONTRACT : 0
4. CUMULATIVE HOURS USED UP TO 26 MAY 1976 : 322 h 20 (TOTAL COL. 3)
5. SPECIAL CONDITIONS :
6. ACTUAL HOURS BILLED AFTER 0 HOURS : 322 h 20
7. EXCHANGE RATE : 1 A.U. = 0.571 (LSG)
8. OTHER COMMENTS :TO BE SENT TO

** NOTE TO FILE 2 : THE NIC ELEMENT OF AU 10 - FOR CAC FILE IS DUE TO ESRO.

PLEASE PAY TO :

Appendix 6

Example of computer generated monthly invoice prepared for ESA RECON dial-up user.

```
ESA -- SPACE DOCUMENTATION SERVICE  -  TOTAL FILE ACCESS FOR RECON USE
----------------------------------     PERIOD: 3 MAY 1976 - 28 MAY 1976
TERMINAL NO. 403-0                     REFERENCE NO.:    76/ACC
                                       DATE        : 1976/ 6/31
```

ATTN: DTI CONTRACT VIA RTC

1		2	3	4	5	6
FILES NO.	DESCRIPTION	USED TIME	CUMULATIVES	OTHERS	FEE PER HOUR IN A.U.	TOTAL IN A.U.
1	N A S A	4 h 20	8 h 11		32	138.67
2	CHEMICAL ABSTR.	0 h 17	1 h 17		46	13.03
3	METADEX	0 h 3	3 h 42		38	1.90
4	COMPENDEX	0 h 12	0 h 35		44	8.80
6	NTIS - GRA	0 h 15	1 h 19		38	9.50
7	USAE - NSA	0 h 36	0 h 36		49	29.40
8	INSPEC	0 h 9	2 h 52		40	6.00
11	ENVIR. SC. IND.	0 h 0	0 h 40		44	0.0
		5 h 52	19 h 12		TOTAL :	207.30 (A.U.)
						140.05 (LSG)

```
EXTRA FEE FOR RTC         6 AU/HR     35.20 (A.U.)

EXTRA FEE FOR FILE 2     10 AU/HR      2.83 (A.U.)
```

N O T E S :

1. CONTRACT NO. : AGR.MENT
2. DURATION OF CONTRACT : FROM 76- 3-25 TO 0- 0- 0
3. HOURS FORESEEN IN CONTRACT : 0
4. CUMULATIVE HOURS USED UP TO 28 MAY 1976 : 19 h 12 (TOTAL COL. 3)
5. SPECIAL CONDITIONS :
6. ACTUAL HOURS BILLED AFTER 0 HOURS : 19 h 12
7. EXCHANGE RATE : 1 A.U. = 0.571 (LSG)
8. OTHER COMMENTS :

** NOTE TO FILE 2 : THE NIC ELEMENT OF AU 10 - FOR CAC FILE IS DUE TO ESRO.

PLEASE PAY TO :

Appendix 7

Example of monthly statement of royalties, or use charges, incurred by ESA RECON user.

```
ESA -- SPACE DOCUMENTATION SERVICE  -  TOTAL FILE UTILISATION FOR RECON USE
----------------------------------        PERIOD: 3 MAY 1976 - 26 MAY 1976
TERMINAL NO. 117                         REFERENCE NO.:      76/ROY
                                         DATE          : 1976/ 6/31
```

ATTN:

1		2	3	4	5	6
FILES NO.	DESCRIPTION	USED TIME			FEE PER HOUR IN $	TOTAL IN $
1	N A S A	13 h 26			0	0.0
2	CHEMICAL ABSTR.	3 h 4			4	12.27
3	METADEX	0 h 58			2	1.93
4	COMPENDEX	1 h 30			3	4.50
6	NTIS - GRA	7 h 15			2	14.50
7	USAE - NSA	0 h 10			0	0.0
8	INSPEC	3 h 52			10	38.67
9	W. A. A.	0 h 35			0	0.0
11	ENVIR. SC. IND.	0 h 14			0	0.0
		31 h 4			TOTAL :	71.87

N O T E S :

1. CONTRACT NO. : 13/73
2. DURATION OF CONTRACT : FROM 75-11-15 TO 76-11-14
3. SPECIAL CONDITIONS :
4. OTHER COMMENTS :TO BE SENT TO

PLEASE PAY IN US$ TO:

ACCOUNT NO. 013.50.11
ESA/SDS
BANCO DI ROMA
FILIALE DI ROMA
ROME - ITALY

UNE BANQUE DE DONNEES DE BIOMETRIE HUMAINE

A.M. COBLENTZ - Maître de Conférences à l'Université René Descartes (PARIS V)

Directeur du Laboratoire d'Anthropologie et d'Ecologie Humaine.

1 - GENERALITES : L'INFORMATION BIOMETRIQUE ET LA CONCEPTION DES EQUIPEMENTS -

La conception et l'élaboration, dans tous les domaines d'activités, d'équipements et de matériels de plus en plus complexes imposent, dès les tous premiers stades des études, une réflexion sur l'organisation ergonomique de ces équipements. Celle-ci doit tenir compte des caractéristiques des populations d'utilisateurs appelés à les servir ou à en utiliser les services, à des échéances connues, plus ou moins lointaines.

De plus, et il s'agit là d'une évolution générale, les équipements et tout particulièrement certains postes de conduite deviennent de plus en plus complexes. Bien souvent ils doivent être utilisés dans des conditions de charge de travail telles pour l'utilisateur, que si les études de relations *hommes-machines* ont été peu ou mal prises en compte dès la *conception*, cet utilisateur n'en obtiendra jamais le maximum d'efficacité *au moment critique*. Les conséquences bien entendu dépassent largement la simple notion d'utilisation, puisqu'elles peuvent concerner la sécurité d'un homme ou celle d'un équipage et compromettre le succès de la mission.

Par ailleurs, ces mêmes équipements sont très souvent destinés à la commercialisation et à l'exportation et de ce fait appelés à être servis dans des conditions qui, fréquemment, ne sont plus celles habituellement rencontrées dans le pays dans lequel ils ont été conçus. Il en résulte que, même s'il le souhaite, le concepteur du matériel ou de l'équipement rencontre des difficultés à tenir compte des caractéristiques biométriques de populations des utilisateurs potentiels et surtout des aspects de *variabilité* de cette biométrie et cela faute d'information.

Pour les bureaux d'études, les *laboratoires*, les services commerciaux, les difficultés à prendre en compte ces ensembles de paramètres biométriques, lors de la conception des équipements, peuvent se situer à différents niveaux :

- en premier lieu, la complexité des problèmes soulevés par l'harmonisation des impératifs biométriques et des contraintes techniques, particulièrement au niveau de la conception de poste, apparaît encore mal maîtrisée au niveau des bureaux d'études, encore que cette préoccupation s'y développe de plus en plus,
- en second lieu, avant la création d'une Banque de Données de Biométrie, les bureaux d'études ou les laboratoires, n'accèdaient que difficilement aux informations biométriques nécessaires à la solution des problèmes abordés,
- enfin, même en possession de ces informations, l'ingénieur de bureau d'étude ou le chercheur de laboratoire observent généralement une inadéquation entre les données qu'ils auront recueillies *sous* forme de moyennes, écarts-types, percentilages, relatives à des groupes humains particuliers et les besoins spécifiques de l'étude dont ils ont la charge.

2 - SOURCES D'INFORMATIONS DISPONIBLES ET INTERET D'UNE BANQUE DE DONNEES -

On ne dispose généralement que de deux sources d'informations : les publications antérieures et les études spécifiques.

2.1 - Publications scientifiques ou rapports techniques -

Les publications antérieures, scientifiques ou techniques sont relativement accessibles, quoique peu de services de documentation possèdent une bibliographie convenable et complète sur le sujet. Un des principaux inconvénients de ce type de documentation provient du fait que les échantillons de populations mesurés n'ont souvent que peu de rapport avec les utilisateurs potentiels d'un matériel mis à l'étude. Généralement les données recueillies pour des enquêtes antérieures l'ont été dans un but différent et leur choix ne répond généralement pas aux besoins. Il devient donc nécessaire d'envisager des corrections et des adaptations.

Le bureau d'étude qui travaille à l'élaboration d'un véhicule de transport ou, de façon beaucoup plus contraignante, sur un poste de pilotage d'avion a peu de chance de trouver les dimensions *taille-assis, longueur fesse-genou,* dans lesquelles est pris en considération l'encombrement de l'équipement individuel. On connaît beaucoup mieux la stature d'un homme dévêtu que celle du même homme chaussé et complètement équipé.

La dimension la plus intéressante à connaître pour évaluer l'implantation d'une commande ne figure dans aucune étude préexistante et l'on en est réduit à faire des évaluations hasardeuses lorsqu'elles ne sont pas proposées par des spécialistes.

Enfin, les résultats sont toujours publiés sous forme de paramètres statistiques : moyennes, écarts-types, maximum et minimum, parfois le percentilage ou le coefficient de variation et très rarement la matrice d'intercorrélations.

A partir de ces données, il est exclu actuellement, pour des non biométriciens, de se livrer au moindre calcul nouveau sur les dimensions sans prendre le risque de se livrer à des approximations abusives. Il suffit à cet égard de prendre comme exemple le problème de la définition d'un modèle de population choisi par le percentilage. Quelles sont les valeurs de variabilité (percentilage minimum et maximum) des dimensions corporelles segmentaires d'un échantillon de population choisi à 5 % pour la variable *stature ?* La présentation des résultats laisse parfois croire aux utilisateurs qu'il suffit de faire la somme des éléments segmentaires qui composent la stature, choisis individuellement à 5 % pour retrouver la population stature 5 % en les additionnant. On en arrive ainsi à la situation paradoxale qui consiste à aggraver *les contraintes liées à la Biométrie !*

2.2 - Etudes anthropométriques spécifiques -

La deuxième source de données dont on dispose est constituée par des études *spécifiques* ou enquêtes directes adaptées au problème posé ou à un problème très voisin. Dans ce cas, les mesures adéquates et les calculs souhaités sont effectués sur un échantillon représentatif de la population. Le coût d'une telle étude peut parfois paraître relativement élevé mais bien souvent lorsqu'il s'agit d'un matériel relativement sophistiqué sa réalisation s'avère parfaitement justifiée.

2.3 - Evolution des caractéristiques biométriques au cours du temps -

Une enquête ne fournit généralement des données que sur la morphologie actuelle d'une population. Une telle information ne satisfait que très partiellement la plupart des concepteurs dont le matériel à l'étude ne sera mis en service que cinq ans plus tard, puis utilisé pendant une période que l'on peut évaluer entre quinze et vingt-cinq ans. On doit tenir compte dans toute la mesure du possible des modifications de la morphologie et si on le peut, de l'incidence des équipements individuels nouveaux susceptibles d'être mis à la disposition des utilisateurs.

3 - LA BANQUE DE DONNEES DE BIOMETRIE HUMAINE DANS L'ABORD DES PROBLEMES DE CONCEPTION D'EQUIPEMENT -

3.1 - Principe de conception de la Banque de Données de Biométrie Humaine -

Les difficultés que rencontre le responsable d'étude et de conception d'équipement sont très fréquemment confiées aux laboratoires spécialisés, et tout particulièrement aux laboratoires d'Anthropologie.

En général, lorsque des problèmes ou des questions sont soumis à ces laboratoires, qui assez souvent n'ont dans leurs archives qu'un capital insuffisant d'information pour apporter des réponses complètes, ils se trouvent confrontés à plusieurs choix:

- parfois ils répondent qu'il leur est impossible de fournir les renseignements demandés dans l'immédiat, car l'exploitation des données existantes correspondant à cette demande n'a pas été effectuée dans un sens que les questions posées requièrent,
- parfois, ils reprennent les données de base d'une enquête dont les fiches de mesure, les cartes perforées ou les bandes magnétiques sont stockées depuis plusieurs années en archives : une telle opération entraîne très souvent un délai de présentation des résultats plus ou moins considérable en relation avec l'activité générale dudit laboratoire,
- parfois enfin, pour reconstituer le modèle de population envisagé, le laboratoire interrogé doit consulter les données de différentes enquêtes antérieures plus ou moins éparpillées. Il se révèle plus rapide souvent de refaire l'enquête sur le terrain, ce qui présente en outre l'avantage d'orienter le travail sur la question précisément posée.

Pour pallier les difficultés qui viennent d'être évoquées, pour rendre plus aisée, plus rapide, plus adaptée, la prise en compte des données anthropométriques dans la conception des équipements, le Laboratoire d'Anthropologie et d'Ecologie Humaine de l'Université René Descartes (Paris V), avec le soutien de la D.R.M.E., a créé une *Banque de Données Internationale de Biométrie Humaine.*

Le principe d'une telle Banque de Données de Biométrie repose sur le rassemblement et la *réutilisation* des mesures anthropométriques individuelles recueillies depuis des décennies sur de très nombreuses populations du monde.

Il s'agit donc de réunir un capital composé du plus grand nombre possible de données disponibles afin de les mettre à la disposition des laboratoires, des bureaux d'études, ou des chercheurs.

Une exploitation rationnelle et efficace de ce fond de données, d'origine très diverse, nécessite avant tout une définition précise des mesures effectuées et le recueil de toutes les caractéristiques explicites ou implicites des populations étudiées.

Les données doivent ensuite être enregistrées sous une forme comprimée, de façon à réduire les temps d'exploitation et les coûts de traitement des interrogations. Grâce à cela on peut envisager des échanges conversationnels avec la Banque.

Quelques exemples vont permettre de mettre en évidence les types de problèmes ou de questions auxquels la Banque de Données peut contribuer à apporter des réponses.

3.2 - Exemples du rôle de la Banque de Données dans l'abord de problèmes biométriques -

3.2.1 - Conception de poste et définition biométrique de la population utilisatrice :

Lors de la conception d'un poste de pilotage, par exemple, il apparaît indispensable, afin de satisfaire le plus grand nombre d'utilisateurs possible, de définir les impératifs *d'habitabilité et d'atteignabilité* des éléments de poste grâce aux mesures biométriques les plus appropriées. Il conviendra ensuite de connaître leur répartition statistique et de choisir à l'intérieur de cette répartition les bornes minimales et maximales pouvant localiser au mieux cette population d'utilisateurs. En d'autres termes ces limites sont retenues en fonction du pourcentage de la population que l'on désire satisfaire.

Il existe donc à ce niveau, avant toute approche expérimentale un ensemble de questions qui se posent :

- quelle est la structure générale de la population utilisatrice *actuelle* et *prévisible* selon les échéances connues de mise en service et d'utilisation du matériel ? En particulier le niveau socioculturel des utilisateurs risque-t-il de se modifier dans les deux décennies suivantes ?
- quelle doit être la constitution précise de la population expérimentale qui participera en laboratoire, sur maquette, à la définition des zones d'activité, distance d'atteinte, etc...?

Dans le contexte de la Banque de Données de Biométrie telle que nous l'avons conçue et définie, l'exploitation systématique du capital de données individuelles ouvre de nouvelles possibilités.

Il est devenu possible de calculer, individu par individu, les valeurs les plus probables de dimensions non mesurées expérimentalement. En effet, si les dimensions segmentaires de l'homme sont susceptibles de variations considérables entre ethnies, groupes professionnels, classes d'âges, etc..., les écarts-types et les corrélations demeurent au contraire d'une grande constance.

Avec une très bonne approximation, il devient possible de reconstituer à partir d'un petit groupe, bien étudié sur le plan biométrique, le profil morphologique complet d'une population représentative des utilisateurs d'un matériel, même si à l'origine on ne possède que quelques mensurations de cette population, à la limite le poids et la taille uniquement.

Il devient également possible de regrouper des enquêtes d'origine diverse pour lesquelles les techniques de mesure sont comparables. On obtient alors des échantillons d'effectifs plus nombreux ou bien, ce qui apparaît tout aussi avantageux, il se révèle possible de recréer, à partir de données individuelles, des sous-échantillons homogènes de groupes humains particuliers (selon des critères géographiques, socioprofessionnels) dont les effectifs deviennent également suffisamment importants pour être pris en considération dans les études ergonomiques.

On conçoit en définitive que l'utilisation de ce capital de données biométriques, provenant de multiples enquêtes, donc de multiples populations, soit extrêmement fructueuse. Elle permet de tenir compte avec la plus grande précision de la variabilité des populations appelées à se servir d'équipements ou de matériels, car compte tenu de l'évolution actuelle et des idées en ce domaine, la conception des équipements quels qu'ils soient, doit intégrer beaucoup plus la notion biométrique et biomécanique.

En effet, à quoi servirait un matériel ou un équipement très sophistiqué, donc très coûteux, si l'utilisateur, placé en situation opérationnelle, comportant fréquemment une charge physique ou perceptive très élevée, ne pouvait obtenir le maximum de performance ?

En outre, le temps de réponse aux questions posées se révèle très bref pour les questions simples et sans commune mesure avec la réalisation d'une étude pour les problèmes plus complexes.

3.2.2 - Vues prospectives : Prévision d'une évolution morphologique

L'intégration dans la Banque de Données d'enquêtes successives étalées dans le temps, mais réalisées sur une même population autorise des extrapolations sans risque et des prévisions concernant l'évolution morphologique de cette population. Ceci appa-

raît d'autant plus important que l'on s'attache à l'étude d'équipements qui sont conçus de 5 à 10 ans avant leur mise en place et demeureront en service pendant 15 à 20 ans.

On sait combien la morphologie (souvent représentée dans la littérature par la seule stature, alors que les dimensions transversales varient également) évolue dans la plupart des populations. Il faut remarquer à cet égard, que l'accroissement de la taille est lié à deux phénomènes, une augmentation réelle de la taille de certaines catégories (phénomènes d'action sur le développement et la croissance) mais aussi une diminution du nombre de sujets de petite taille. On donne à cette évolution une interprétation essentiellement biologique associée étroitement à des phénomènes sociaux et culturels.

Grâce aux données morphologiques, il devient possible d'obtenir une estimation relativement précise (suffisamment précise pour la finalité recherchée) de l'évolution prévisible au cours du temps non seulement de la taille mais encore des autres dimensions segmentaires. Cette estimation ne peut bien entendu se faire qu'en fonction des données sociodémographiques et socioculturelles relatives à la population étudiée.

On conçoit dès lors toute l'efficacité de l'outil que représente la Banque de Données de Biométrie entre les mains d'équipes compétentes de laboratoires ou de bureaux d'études. Il devient possible de réaliser les meilleurs compromis au niveau des *interfaces Homme-Equipement*, sachant qu'en mettant à l'étude un projet de poste de pilotage, un véhicule de transport en commun, une machine outil ou un projet d'architecture, on peut d'emblée intégrer les paramètres biométriques des futurs utilisateurs, à des échéances d'au moins 15 ans, avec un risque d'erreurs pratiquement nul.

En définitive, il apparaît clairement que la liste des besoins et des utilisateurs potentiels dépasse largement le contexte des quelques exemples qui viennent d'être évoqués. Qu'il s'agisse des domaines de l'équipement individuel, vêtements, gants, masques de protection, des équipements collectifs, transports, machines, véhicules terrestres ou aériens, de l'architecture de l'urbanisme avec les multiples aspects que recouvrent ces domaines, la biométrie intervient généralement en amont de toute réalisation, au niveau de la conception ; sa prise en compte acquiert d'autant plus de valeur, que les équipements conçus sont plus sophistiqués, plus coûteux et doivent être servis dans ces éventualités par des personnels très strictement sélectionnés.

4 - DESCRIPTION SCHEMATIQUE DE LA STRUCTURE INFORMATIQUE DE LA BANQUE -

La réalisation du *logiciel* a été guidée par deux principes de base :

- *La réduction au minimum indispensable* du volume d'informations transférées au cours d'une interrogation.
- *La séparation très nette du processus d'extraction* des processus de traitement statistique des données et le développement d'une gamme étendue de modalités d'extraction.

4.1 - Réduction du volume d'informations transférées -

La réduction au minimum de volume d'informations transférées a été obtenue grâce aux dispositions suivantes :

- chacun des lots de valeurs correspondant aux diverses rubriques d'une enquête est rendu accessible directement,
- un algorithme simple de compression de données permet de réduire au minimum la place occupée par les lots de valeurs.

4.2 - La séparation des processus d'extraction -

Du point de vue du logiciel, une Banque de Données peut se présenter comme un mode *d'extraction généralisé* des informations. Il s'agit donc d'un système qui permet le traitement efficace de *pratiquement toutes* les questions possibles dans le domaine de la Biométrie et de la Biomécanique.

La séparation du processus d'extraction du processus de traitement est la première condition d'une telle efficacité. L'adjonction d'un traitement nouveau se limite ici à des calculs nouveaux sur des données extraites. *L'inventaire des possibilités d'extraction* est la deuxième condition d'efficacité. S'il n'est pas possible de prévoir d'emblée, toutes les modalités d'extraction, il faut au moins en prévoir un grand nombre et faire en sorte que le système d'extraction soit ouvert, c'est-à-dire que l'adjonction d'une nouvelle possibilité d'extraction soit aisée.

4.3 - Organisation des programmes -

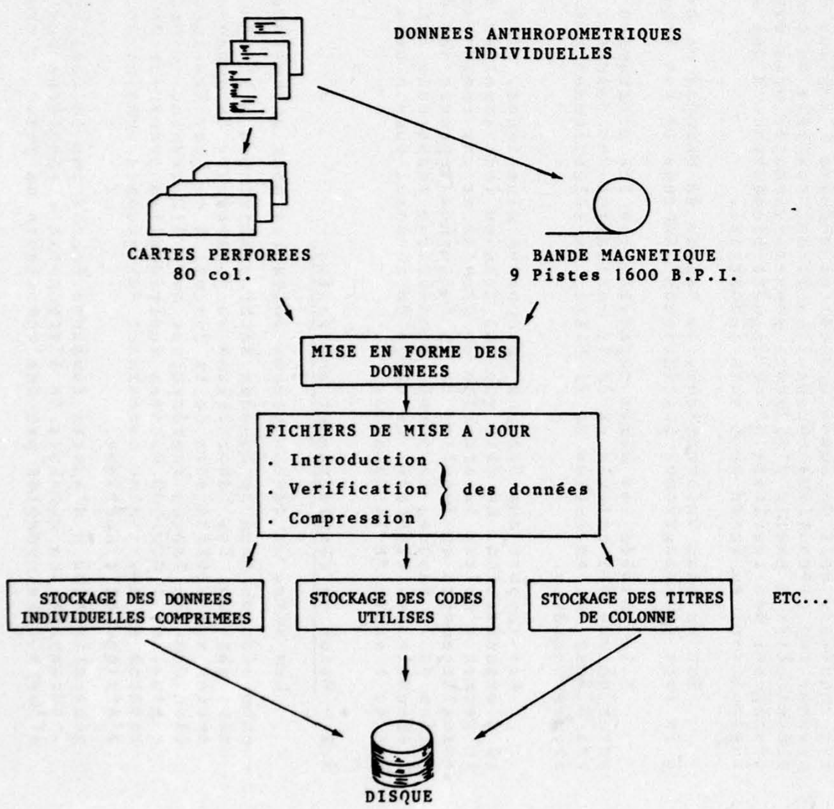
Le logiciel comprend deux groupes de programmes. (Cf. schémas suivants).

4.3.1 - Le programme d'interrogation :

Ce programme assure trois fonctions principales :

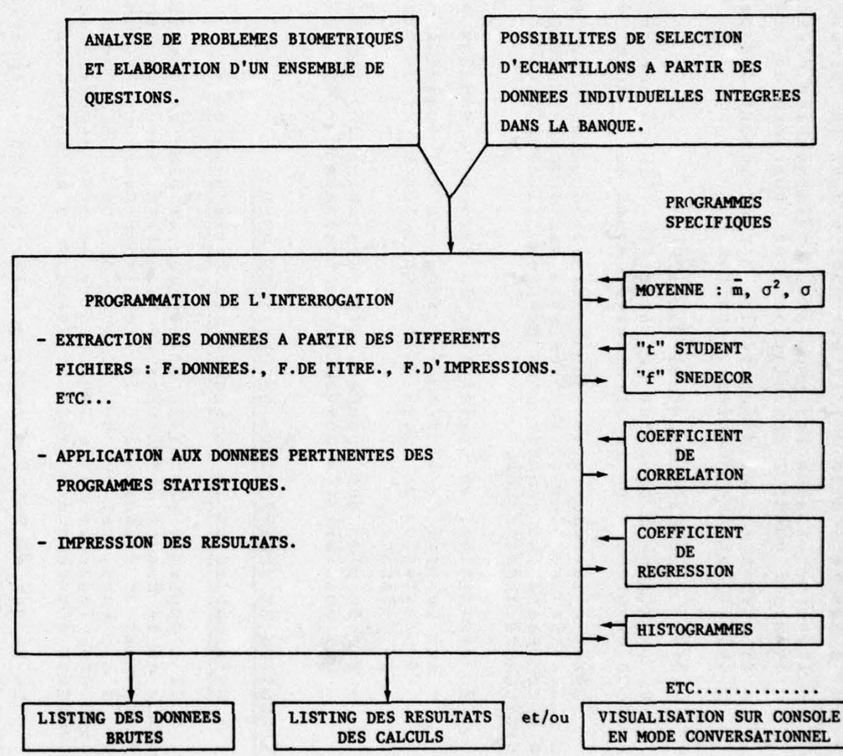
STRUCTURE DE LA BANQUE DE DONNEES DE BIOMETRIE

I - INTRODUCTION DE FICHIERS DE DONNEES -



STRUCTURE DE LA BANQUE DE DONNEES DE BIOMETRIE

II - INTERROGATION DE LA BANQUE -



- il assure l'aide à l'interrogation dans la formulation des questions,
- il permet toutes les modalités envisageables d'extraction des données et éventuellement l'association à un programme de traitement,
- il permet enfin, d'introduire à tout moment des possibilités nouvelles de calcul sur les données extraites.

4.3.2 - Le programme de mise à jour :

Ce programme a pour but de permettre l'introduction à tout moment des données d'origine diverse, dans la structure commune du logiciel de la Banque.

Avant d'introduire les données individuelles d'une enquête dans la Banque, il est nécessaire de décrire au préalable ses divers constituants, sur cartes. On distingue à cet égard des *constituants généraux* (indépendant d'une enquête donnée) et des *constituants spécifiques*.

En définitive, le logiciel de la Banque de Données présente un double avantage :

- sur le plan de la *programmation*, un tel logiciel permet, avec un maximum d'adaptation de faire face à tous les types de traitements susceptibles de se présenter,
- sur le plan des données, l'accumulation des informations comparables ouvre des possibilités nouvelles de traitement.

5 - CONCLUSIONS ET PERSPECTIVES DE DEVELOPPEMENT -

5.1 - Originalité de la Banque de Biométrie Humaine -

Il n'apparaît pas utile de développer plus en détail les possibilités d'intervention de la Banque de Données de Biométrie dans tous les domaines de conception d'équipement et d'adaptation de l'environnement aux besoins des utilisateurs. Sa création, d'ailleurs, est née de l'importance de la demande en ce domaine et de l'impossibilité dans laquelle notre Laboratoire et très certainement de nombreux autres se trouvaient d'y répondre faute de disposer d'un instrument de ce type.

La Banque de Données telle qu'elle est conçue apparaît comme un outil d'utilisation très souple et complet, capable de traiter au *moindre coût* de nombreux problèmes d'ergonomie dans lesquels interviennent les aspects de dimensions et de variabilité morphologiques. Il ne peut y avoir d'ergonomie de conception sans biométrie.

Alors que traditionnellement chaque problème posé se traduit par l'étude anthropométrique spécifique d'un échantillon plus ou moins représentatif d'une population donnée, la Banque de Données permet, dans la plupart des cas, soit d'éviter de nouvelles enquêtes, soit de considérablement réduire l'ampleur de l'enquête nécessaire. En prenant les précautions nécessaires il est possible de constituer des échantillons représentatifs à partir d'éléments sous-représentés pris dans plusieurs enquêtes, de transposer des résultats d'une enquête biométrique à une autre avec un risque d'erreur insignifiant eu égard au besoin industriel.

Sur un plan informatique, la Banque de Données de Biométrie Humaine se distingue à la fois des programmes statistiques généraux et des Banques de Données classiques.

Elle possède les mêmes capacités que les programmes statistiques généraux du point de vue des calculs et de la validation des données. Elle permet en outre de mettre à jour des ensembles de fichiers, de sélectionner des sous-ensembles à travers plusieurs enquêtes.

Par rapport aux Banques de Données classiques, elle se distingue par la simplicité d'organisation. Le fait que les données individuelles soient introduites par enquête permet d'éviter la création de *groupes* et par conséquent la hiérarchisation des enregistrements en fonction de la fréquence supposée des demandes. Lors d'une interrogation de la Banque, la donnée biométrique recherchée est extraite par exploration de l'ensemble des fichiers. Le temps de réponse, compte tenu des programmes mis en place, est de l'ordre de la seconde.

5.2 - Voies de développement envisagées -

Les voies de développement suivantes sont actuellement envisagées :

- extension du fond de données anthropométriques à l'aide de nouvelles enquêtes fournies par tous les laboratoires ou possesseurs de données biométriques désirant s'associer aux utilisateurs de la Banque. A ce jour les informations en cours d'introduction dans la Banque représentent avec 310 mesures anthropométriques différentes, un capital de 1.800.000 données individuelles provenant de différents laboratoires européens et américains concernant des groupes humains nombreux et variés d'Europe, d'Afrique et d'Amérique,
- généralisation à d'autres domaines fondés sur le développement d'enquêtes régulières intéressant les domaines de l'ergonomie : réactions aux vibrations, variabilité des efforts développables par des opérateurs au poste de travail, etc...

- développement de la programmation directe par les utilisateurs. D'ores et déjà un certain nombre de calculs simples peuvent être demandés directement sur le mode conversationnel. En compliquant les possibilités de calcul et en introduisant les conditions logiques, un utilisateur doit pouvoir traiter directement, sans grand apprentissage, des problèmes relativement complexes,
- la liaison avec une console de visualisation est actuellement envisagée. En effet il est apparu au cours d'expériences de simulation que de nombreux problèmes de biométrie et d'ergonomie - conception d'engins et de postes de conduite ou de commandes morphotypes de structures anatomiques (mains, face, pied, etc...- pouvaient être résolus avec un gain de temps considérable par une simulation avec visualisation.

Dans certains problèmes, le jeu de questions et de réponses immédiates en *mode conversationnel* facilite la progression et permet, par essais successifs, d'aboutir à une solution satisfaisante.

De même, la possibilité de modéliser sur écran les différentes éventualités de choix qui se présentent généralement, dans toutes études de conception d'équipement, en raison des compromis permanents recherchés entre les contraintes techniques et les aspects de la variabilité biologique, apparaît dès à présent une voie fructueuse qu'il convient d'exploiter.

A HUMAN BIOMETRY DATA BANK

by

A.M.Coblentz, Senior Lecturer
at the Université René Descartes, Paris V,
and Director of the Anthropology
and Human Ecology Laboratory

1. GENERAL: BIOMETRIC DATA AND EQUIPMENT DESIGN

In all spheres of activity the design and development of systems and equipment of ever increasing complexity make it essential to give consideration at the very earliest design stages to the ergonomic layout of such equipment. This must take into account the characteristics of the populations of users who will have to operate the equipment or use its services at known dates in the fairly distant future.

In addition, and this is a general trend, all items of equipment, and particularly certain types of vehicle driving compartments and aircraft cockpits are becoming increasingly more complex. Quite often they have to be operated in workload conditions which are such that, if the man-machine relation studies have not been taken into sufficient account, or have been incorrectly taken into account at the design stage, the user will never obtain maximum efficiency from the equipment at the critical moment. The resulting consequences will of course go far beyond the mere concept of use, since they may affect the safety of a man or of a crew and compromise the success of a mission.

Furthermore, the same equipment is very often intended for marketing and export and is thus required to be operated in conditions which are frequently not the same as those normally found in the country in which it was designed. But, even if the equipment or system designer wanted to take account of the biometric characteristics of populations of potential users, and particularly of the variability aspects of such biometry, he is finding it difficult to do so, simply because the relevant facts are not available.

In the case of design offices, laboratories and commercial departments of firms, the difficulty of taking such sets of biometric parameters into account at the design stage of the equipment arises at various levels:

- firstly, the complexity of the problems involved in harmonising biometric requirements and technical constraints, particularly at the design level, still appears to have been inadequately overcome by design offices, even though this is a matter which is of increasing concern;
- secondly, before a Biometry Data Bank was set up, design offices and laboratories had great difficulty in gaining access to the biometric data required for solving problems under investigation;
- finally, even when he obtains these data, the design engineer or the laboratory researcher generally finds a great gap between the data given to him in the form of averages, standard deviations or percentiles relating to particular human groups, and the specific requirements of the design for which he is responsible.

2. AVAILABLE SOURCES OF INFORMATION AND THE ADVANTAGES OF A DATA BANK

Only two sources of information are generally available: previous publications and specific studies.

2.1 Scientific Publications or Technical Reports

Previous scientific or technical publications are relatively accessible, although few documentation services have a suitable and comprehensive bibliography on this subject. One of the main disadvantages of this type of documentation arises from the fact that the population samples measured often bear very little relationship to potential users of an item of equipment currently under design. The data collected for previous surveys have generally been for a different purpose and their selection does not normally fit specific requirements, thus leading to the need for corrections and amendments.

The design office working on the design of a public transport vehicle or, something which imposes many more constraints, the cockpit of an aircraft, has very little chance of finding the dimensions relating to sitting height or buttock to knee length within which the overall size of the equipment has to be designed. Much more information is available on the height of a man without his clothes than on the height of the same man with his shoes on and wearing all his equipment.

The most useful measurement to know when assessing the location of a control never appears in any existing study and one is reduced to making assessments which may be dangerous if they have not been made by specialists.

Finally, the results are always published in the form of statistical parameters: averages, standard deviation,

maximum and minimum values, occasionally percentiles or the coefficient of variation, and very rarely, the intercorrelation matrix.

At the present time non biometricians are excluded from making the least new calculation about dimensions using the above data, for fear of the risk of arriving at a wrong approximation. An obvious example of this is the problem of defining a sample population selected on the basis of percentile values. What are the variability values (minimum and maximum percentiles) of the segmental physical measurements of a 5% population sample in the case of the 'height' variable? The results are often presented in a form which leads users of the data to believe that all that is necessary is to add together the various segmental data constituting 'height', selected on a 5% basis, in order to find the height of 5% of the population. We thus arrive at the paradoxical situation of aggravating the constraints associated with biometry!

2.2 Specific Anthropometric Studies

The second available source of information consists in specific studies or direct surveys catering for the particular problem involved or one very much akin to it. In such case, suitable measurements and the desired calculations are carried out on a representative sample of the population. The cost of such a survey may sometimes seem relatively high but very often proves fully justified if the equipment concerned is relatively sophisticated.

2.3 Changes in Biometric Characteristics over a Period of Time

A survey will generally provide data only on the current morphology of a population. But such information can only partially meet the requirements of the majority of designers who are working on equipment which will not come into service until five years later, and then be used for a period estimated at between fifteen and twenty-five years. Maximum account should be taken of changes in morphology, together with, if possible the number of new individual equipments likely to be available to users.

3. THE HUMAN BIOMETRY DATA BANK AND EQUIPMENT DESIGN PROBLEMS

3.1 Basic Principles of the Design of the Human Biometry Data Bank

Any difficulties encountered by equipment designers are very often referred to specialist laboratories, and particularly to anthropology laboratories.

Generally speaking, because the stock of data held in their files is very often insufficient to enable them to give complete answers to the problems or questions put to them, these laboratories find themselves having to make one of several choices when replying:

- sometimes they reply that they cannot immediately supply the information requested as the existing data relating to the request have not been analysed from the particular aspect required by the questions asked;
- sometimes they go back to the basic data obtained during a survey, the relevant forms, punched cards or magnetic tapes for which have been filed away for several years: this operation very often means that the results may take some time to produce, depending on the general workload of the laboratory;
- finally, on some occasions, in order to reconstitute the sample population being considered, the laboratory to which the matter has been referred has to consult data obtained from previous more or less scattered surveys. It often turns out to be quicker to conduct the survey again on the spot, and this has the added advantage that the investigation can be directed to the specific question raised.

To overcome the difficulties just mentioned, and to make it easier and quicker to take account of anthropometric data when designing equipment, and to do so more appropriately, the Anthropology and Human Ecology Laboratory of the Université René Descartes (Paris, V), with the support of the Direction des Recherches et Moyens d'Essais (Directorate of Research and Test Facilities) has set up an International Human Biometry Data Bank.

The basis of a data bank of this kind is the collection and re-use of individual anthropometric measurements gathered over several decades on a large number of world populations.

It is therefore a matter of assembling a stock of the largest possible number of available data items and placing them at the disposal of laboratories, design offices and researchers.

A primary requirement for rational and efficient application of this data base, created from various sources, is a precise statement of the measurements made, and the collecting together of all the explicit or implicit characteristics of the populations studied.

The data must then be recorded in a compressed form so as to reduce the time required for their application and the cost of processing interrogations. Conversational exchanges with the bank can then be envisaged.

A few examples will illustrate the types of problems or questions for which the data bank can help to provide answers.

3.2 Examples of the Part played by the Data Bank in tackling Biometric Problems

3.2.1 Design of an Item of Equipment and Biometric Definition of the User Population

When a cockpit for example is being designed, it would seem essential, in order to satisfy the largest possible number of users, to define the 'habitability' (or environmental) requirements, and the 'reachability' requirements for the various components in the cockpit, using the most appropriate biometric measurements. It is then helpful to know the statistical distribution of these measurements and to choose, within this distribution, the minimum and maximum limits within which to place this population of users. In other words, these limits are adopted in relation to the percentage of the population whose requirements are to be met.

A number of questions therefore arise at this level before any experimental approach can be made:

– What is the general structure of the present and foreseeable user population according to the known dates of entry into service of the equipment concerned? In particular, is the social and cultural level of the users likely to alter during the next two decades?

– What is to be the exact constitution of the experimental population which will take part, in the laboratory or in a mock-up, in defining the areas of activity, reaching distance etc?

In the context of the Biometry Data Bank as designed and defined by us the systematic use of the stock of individual data items held in store opens up new possibilities.

It is now possible to calculate for each individual person the most probable values of the dimensions which have not been measured experimentally. In fact, although the segmental measurements of a person are likely to vary considerably between ethnic groups, professional groups, age classes etc, standard deviation values and correlations remain on the other hand largely constant.

A small group which has been subjected to a careful biometric study can help to reconstitute the complete morphological profile of a representative population of users of a particular equipment, even though initially only a few measurements in regard to this population were available, at most, weight and height only.

It is also possible to regroup surveys undertaken for various reasons, but for which the measuring techniques are comparable. This will result in larger samples or, and this would appear to have just as many advantages, individual data items can be used to recreate homogeneous sub-samples of particular human groups (based on geographical, social and professional criteria), the members of which are also sufficiently numerous to be taken into consideration for the purpose of ergonomic studies.

It will be realised, finally, that the use of this capital stock of biometric data gathered during a very large number of surveys, and therefore on a very large number of populations, can provide extremely fruitful results. It enables highly accurate account to be taken of the variability of the populations required to use systems and equipment, since in view of present developments and ideas in this field, the design of equipment of whatever kind is bound to incorporate to a much greater extent the concept of biometry and biomechanics.

Indeed, what would be the good of a very sophisticated, and therefore very costly item of equipment if the user, finding himself in operational conditions frequently involving great physical and mental responsibility, were unable to derive maximum efficiency from it?

Furthermore, the time taken to reply to questions is very short in the case of simple questions and without any common measure with the performance of a study for solving more complex problems.

3.2.2 Prospection: Predicting Morphological Changes

As a result of incorporation in the Data Bank of the results of succesive surveys conducted over a period of time, but using the same population, it is possible to extrapolate without risk of error and to predict the morphological development of this population. This would appear to be especially important as we are concerned with work on equipment which is being designed five to ten years before it is due to come into service and which then remains in service for some fifteen to twenty years.

The extent to which morphological changes (often represented in the literature merely by height, whereas the transverse dimensions also vary) occur in the majority of populations is well known. It should be noted in this connection that increased height is associated with two phenomena: an actual increase in the height of certain categories (phenomena affecting development and growth) and also a decrease in the number of subjects who are small in stature. These changes are regarded mainly as biological changes closely dependent on social and cultural phenomena.

Using morphological data it is possible to make a relatively precise estimate (precise enough for the final result desired) of the foreseeable development with time not only of stature but also of the other segmental dimensions. This estimate can of course be made only on the basis of the sociodemographic and sociocultural data relating to the population studied.

The effectiveness of such a tool as the Biometry Data Bank in the hands of competent laboratory or design office teams can therefore well be imagined. It allows them to achieve the best compromises for the man-machine interfaces, knowing that when designing a cockpit or a public transport vehicle, a machine tool or an architectural project, they can incorporate immediately the biometric parameters of those who will be using the equipment at dates at least fifteen years hence, and with practically no risk of error.

Finally, it is clear that the list of requirements and potential users is largely outside the context of the few examples just quoted. Whether we are concerned with equipment for the individual: clothing, gloves, protective masks, or equipment for collective use: means of transport, machines, land vehicles or aircraft, architectural projects for town planning, with the very many aspects of all these various fields, biometry is generally involved before the final development stage, e.g. at the design stage; taking biometry into account is all the more important as the equipment designed becomes more sophisticated and more expensive and for these reasons has to be operated by very carefully selected personnel.

4. DIAGRAMMATIC DESCRIPTION OF THE DATA PROCESSING STRUCTURE OF THE BANK

The creation of the software has been based on two guiding principles:

- reducing to the absolute minimum the volume of data transferred during an interrogation;
- separating the extraction process from the statistical processing operations of the data and developing a comprehensive range of extraction facilities.

4.1 Reducing the Volume of Data Transferred

The volume of data transferred has been reduced to a minimum by means of the following:

- each of the sets of values corresponding to the various headings of a survey can be accessed direct;
- a simple data compression algorithm, makes it possible to reduce to a minimum the space occupied by the sets of values.

4.2 Separating the Extraction Processes

From the software point of view a data bank can be a generalised means of extracting data. The system concerned is therefore one which processes efficiently practically all the questions which could possibly arise in the field of biometry and mechanics.

Separating the extraction process from the processing operations is the first pre-requisite for this kind of efficiency. The addition of further processing is limited here to further calculations based on the data extracted. An inventory of the extraction facilities is the second requirement for efficiency. If it is not possible to provide from the start for the inclusion of all the extraction facilities, provision for at least a large number should be made and the extraction system designed so that it is open-ended, thus enabling a new facility to be added without difficulty.

4.3 Program Organisation

The software comprises two types of programs (see accompanying diagrams).

4.3.1 The Interrogation Program

This program performs the following main functions:

- it assists, during the interrogation, with formulation of the questions;
- it permits the functioning of all the data extraction facilities which can be envisaged and, if necessary, their association with a processing program;
- finally, it makes it possible to introduce at any time further possibilities for calculations based on the data extracted.

4.3.2 The Update Program

This program enables data from various sources to be input at any time into the common structure of the Bank's software.

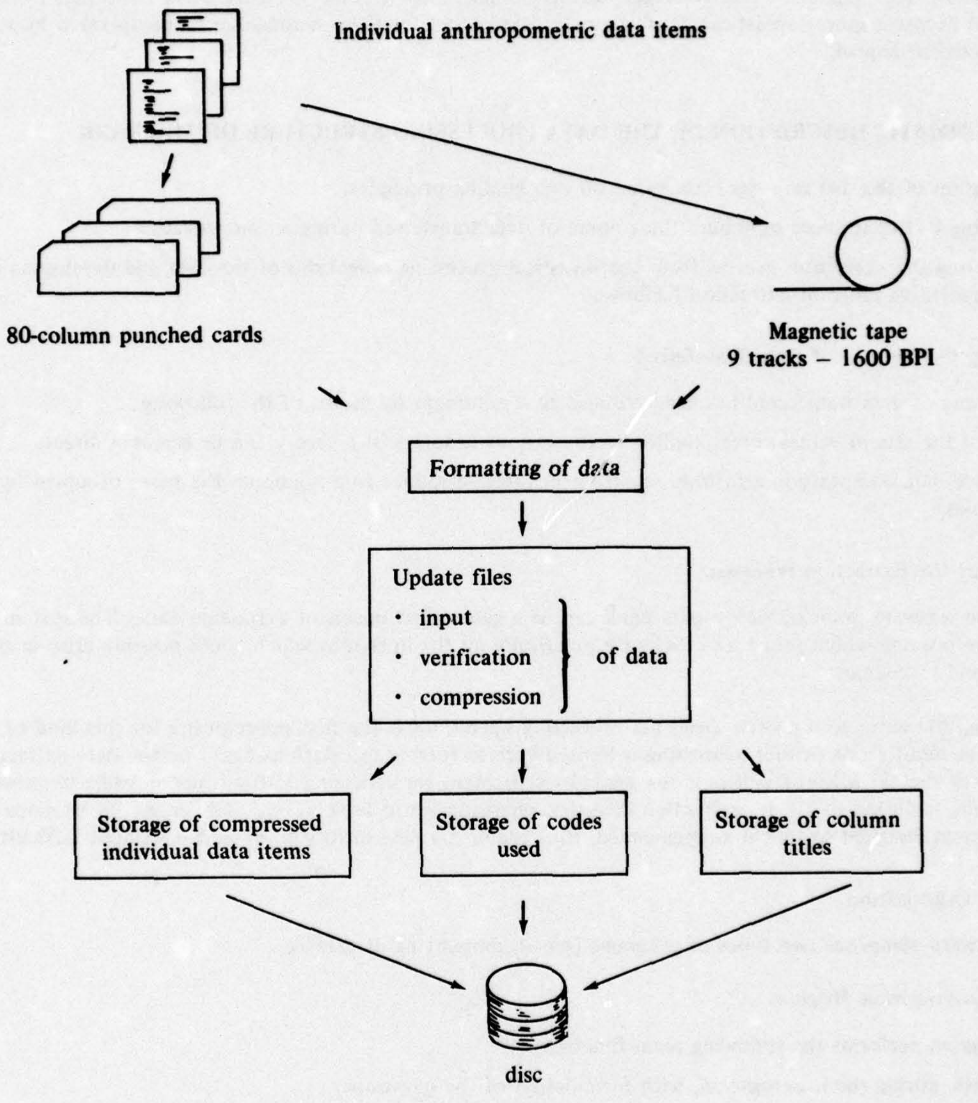
Before the input of the individual data items of a survey into the Bank, the various constituents of the survey must be described beforehand on cards. Here a distinction is made between general constituents (not related to a given survey) and specific constituents.

Finally, the Data Bank software has the following dual advantage:

- from the programming aspect, this type of software makes it possible, with maximum adaptation facilities, to cater for all the types of processing which may arise;
- from the data point of view, the accumulation of comparable data opens up further processing possibilities.

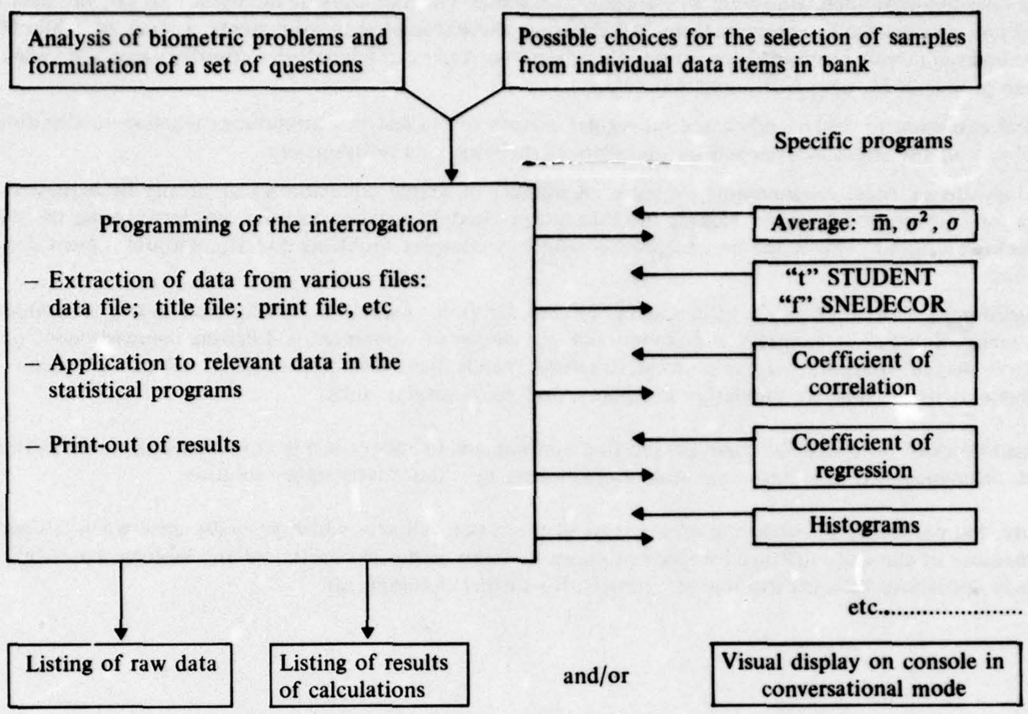
STRUCTURE OF THE BIOMETRY DATA BANK

I. Input of Data Files



STRUCTURE OF THE BIOMETRY DATA BANK

II. Interrogation of the Bank



5. CONCLUSIONS AND DEVELOPMENT PROSPECTS

5.1 Originality of the Human Biometry Data Bank

It would appear unnecessary to go into further detail about the possible ways in which the Biometry Data Bank can be of use in all areas of the design of equipment and making the environment suitable for the requirements of users. Its creation arose moreover from the extent of the demand in this field and the inability of our Laboratory, and most certainly of many others, to cater for this demand because of the lack of a tool of this type.

Because of its particular design features the Data Bank is a very versatile and complete tool, capable of dealing, at a minimum cost, with numerous ergonomic problems involving aspects concerned with morphological measurements and variability. Design ergonomics is impossible without biometry.

Whereas with traditional methods each problems which arises results in the specific anthropometric study of a more or less representative sample of a given population, the Biometry Data Bank makes it possible in the majority of cases either to avoid having to conduct further surveys, or to reduce considerably the scope of any survey which may be necessary. By exercising due care it is possible to constitute representative samples from sub-represented elements taken from several surveys and to transfer the results of one biometric survey to another with an insignificant risk of error, taking into account the industrial requirement in question.

On the data processing level the Human Biometry Data Bank differs both from the general statistical programs and from those of conventional data banks.

It has the same capabilities as the general statistical programs from the computational and validation of data points of view. It can also update sets of files and retrieve sub-sets from several surveys.

Compared with conventional data banks, its distinguishing feature is its simplicity of organisation. The fact that the individual data items are introduced for each survey avoids the creation of groups and consequently avoids having to give a hierarchial structure to the various records according to the assumed frequency of the demands. When the bank is interrogated, the biometric data item being sought is extracted by scanning all the files. The response time, depending on the programs involved, is approximately one second.

5.2 Proposed Lines of Development

The following lines of development are under consideration:

- expanding the anthropometric data base by means of new surveys provided by all the laboratories or those possessing biometric data who wish to associate themselves with the users of the Bank. So far, the data now being put into the Bank represent, from 310 different anthropometric measurements, a stock of 1 800 000 individual data items received from various European and American laboratories covering large and varied human groups in Europe, Africa and America;
- general extension to other fields based on regular surveys of interest to ergonomics: reaction to vibration, variability of the stresses developed by operators at their working positions etc;
- development of direct programming by users. A number of simple calculations can already be requested direct using the conversational mode. Making the calculation capabilities more complex and introducing the logic requirements should enable the user to process relatively complex problems directly, without a great deal of training;
- connection to a visual display unit is now under consideration. Indeed, it became clear during simulated tests that many problems in biometry and ergonomics, e.g. design of equipment and driving compartments, or controls shaped to accord with anatomical structures (hands, face, feet etc), could be solved with a considerable saving of time by simulation techniques and visual display units.

In certain types of problems the game of question and immediate answer in the conversational mode facilitates progress and, after successive trial and error evaluations, finally provides a satisfactory solution.

Similarly, the possibility of modelling on a screen all the various choices which generally arise when designing equipment because of the continual need for compromises between technical constraints and biological variability aspects already appears to be a fruitful line of approach for further development.

THE VIRTUAL-SYSTEM CONCEPT
OF
NETWORKING BIBLIOGRAPHIC INFORMATION SYSTEMS*

J. FRANCIS REINTJES**

Over the past ten years the number of machine-readable bibliographic data bases available to professional communities has increased steadily. The current number is estimated to be in the hundreds and is growing monthly. Many of these are searchable through batch processing of search requests, but there is a steady trend toward the online mode of operation. In general, each commercially available online bibliographic storage and retrieval system offers several data bases to its clientele, with each data base and each storage and retrieval system, having its own set of access procedures and protocols. In the science and technology domain, for example, Lockheed's data bases, numbering approximately 24, are accessible through their DIALOG retrieval system which is implemented on an IBM 360/50 in Palo Alto, California; the dozen or so data bases at System Development Corporation, Santa Monica, California, are on an IBM 370/158 and available through their ORBIT system; the National Library of Medicine has two implementations of its Medline data bases, one at Bethesda, Maryland, the other at the State University of New York at Albany. Since all these systems, as well as many others not cited, have special access requirements and since each data base within each system is nonstandardized, a heavy burden of learning is placed on the end user --- the seeker of information --- in his attempts to engage the system personally. Invariably, in the interest of streamlining the search process and expediting results, a professional intermediary with operational skills in the array of available systems is interposed between host systems and seeker; but even the intermediary is finding it difficult to cope, as the number of data bases and systems continue to proliferate.

It is generally agreed that, from a retrieval effectiveness viewpoint, it is highly desirable to allow the seeker of information to engage the information system himself rather than to have him work through another person. Because of the heterogeneity that presently exists among data bases and systems that contain them, however, placing the seeker directly online is impractical. Two solutions immediately come to mind: 1) standardize all data bases, access procedures and the hardware and software of the host systems to the fullest degree possible; or 2) convey the impression of standardization to the information seeker through use of a computerized interface translator interposed between end users and the systems they wish to access. The interface thus creates a uniform "virtual" system, and it is this single, virtual system that the user engages. This paper reports upon research being undertaken on the virtual-system concept.

The logical structure of a virtual system is shown in Fig. 1. The host retrieval systems with their stored data bases are designated "SYS 1", "SYS 2", and so forth. These may be geographically dispersed nationally or even on a worldwide basis. The end users are U_1, U_2, U_3 ..... U_n. Instead of accessing the retrieval systems directly over communications channels, users work through an interface (we call it the CONIT, standing for <u>CO</u>nnector for <u>N</u>etwork <u>I</u>nformation <u>T</u>ransfer). The objective of CONIT is to make the network of systems and their data bases appear as nearly identical as practicable to end users. CONIT's role is to make the necessary translations of commands and responses and to manage the information flow back and forth between user and the system he is engaging, so that the user gains the impression he is dealing with a network of identical systems.

To accomplish this, CONIT logically has three functions to perform: it interprets user commands for the network and represents individual-system responses to the users in a uniform manner; it manages the many protocols that must be dealt with either locally at the interface, or remotely as information flows to and fro between user and host system; and it provides the necessary translations of commands, messages and protocols between CONIT and the individual host system being accessed.

---

* The work reported here was made possible through grants SIS74-18165 and SIS75-22946 from the National Science Foundation.

** Professor of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

The central issue in configuring a translational interface concerns the degree of "virtualness" that one should expect to attain. Bound up in this issue, in turn, is the amount of computability --- processing and storage --- that one is compelled and willing to commit as overhead in order to achieve a given level of virtualness. From the viewpoint of a seeker of information a virtual system should have several characteristics:

(1) The system should be simple to use and require little or no prior knowledge of procedures on the part of the user. This implies an online step-by-step instructional mode which in essence leads the user "by the hand" through his search session. It also implies a command language that is derived from the user community's native tongue. Ideally, the command language should allow free-form natural-language expressions, but for practical reasons, it will very likely have to be pre-structured.

(2) For those who are expert in the use of the network, a terse mode in which commands are abbreviated and instructions are held to a minimum should be available in order to shorten the time required to make requests and to reduce host-system "chatter."

(3) As a minimum, it should be necessary for the user only to specify the host system and the data base within the host that he wants to engage. No further effort on the part of the user should be required in order to be linked to the data base. In other words, logins and logouts should be automatic. A more sophisticated version of an interface should go even further, since it is really <u>information</u> the user is seeking, not a specific system or data base. Ultimately, the user should be able to begin by making his request for a specific piece of <u>information</u> and then leave it to the interface to select the system and data base(s) where the information is likely to reside. One way to achieve this goal is through use of a <u>Ma</u>ster <u>I</u>ndex and <u>T</u>hesaurus (MAIT) stored at the interface. In principle, the MAIT would consist of a collection of all subject-index terms and perhaps authors associated with all the data bases in the system and the particular systems and data bases that contain the subject terms or authors' names. Upon receiving a subject-search command, the interface would first consult the MAIT to determine where the search should be made and then make the connection to the proper data base of the proper system and execute the command.

(4) Both the minimum and advanced performance requirements in (3) necessitate that all communications connections and system login procedures including dial-ups, hand-shaking procedures and acknowledgements be handled automatically; a similar set of automatic procedures should also prevail for logouts.

(5) Search sessions should be conducted in a single, common-command language that is independent of host-system command language.

The degree of virtualness of system responses that should be built into the network is an open question and not resolvable until some experimentation with bona fide users has been completed. It may not be necessary, for example, to require that easily understood system responses be mapped into a standardized format. Individual system responses, although different from one another, may be easily understood in their original forms.

Another open question that requires study pertains to procedures for extracting the totality of information on a given search request from the various data bases in the various systems comprising the network. Inevitably, there will be redundancies. In these instances the question is: Should duplications be recognized automatically and omitted before presentation to the user, or should recognition of redundancies be left to the user? Implementation of the automatic recognition and combinatorial process is not trivial in the presence of nonstandard formatting.

Still another unresolved matter, as one attempts to get actual users online, is how to design an effective set of online instructions that will lead the user to the information he wants without getting "struck" enroute. A satisfactory approach seems to lie somewhere between the extremes of a highly intelligent automatic instructional mode in which the system is always "in charge" and knows always the next correct action to be taken by the user, and a highly simplified set of rigid instructions which, in effect, provides only one road map through the search session.

To gain a better understanding of some of these issues, an experimental interface has been implemented on a Honeywell 6180 machine, referred to at M.I.T. as "Multics."

The information network included in the experimental setup is shown in Fig. 2. At present the experimental COnnector for Information Transfer (CONIT) contains some, but not all, the features and requirements cited above. The following are salient features that have been implemented to date.

* Five host bibliographic-information systems are accessible through the CONIT. They are Lockheed, Systems Development Corporation, two Medlines (one at Albany, New York, the other at Gaithersburg, Maryland) and the M.I.T. IBM System/370, which contains the Intrex data base.

* Access to all these systems (except the local M.I.T. system) is gained through the Tymshare communications network. Since Multics and the National Library of Medicine's System/370 are in the ARPA network, Medline at Gaithersburg can also be accessed through this network. In addition, there exists a network patch which permits access to the local Intrex data base through a local ARPA Terminal Interface Processor (TIP). See Fig. 2.

* In its current state of development the user must first specify to CONIT which system he wants to access and the data base within the system he wishes to search.

* It is possible to be connected to any system automatically through use of the pick command. Similarly, a data base within the selected system is activated automatically through use of the pick data command. Each command is followed by an argument which defines the system or data base desired.

* Some of the commands for searching a data base, manipulating retrieved information and printing output which have been implemented thus far are these: for making a subject search the find command, followed by an appropriate argument is used. The argument may pertain to document titles, one or more fields of each document record or to a specific named set of documents. The show command, followed by an appropriate argument is used to display one, several or all fields of a document record. Documents retrieved through use of a specific find command may be given a set number for later retrieval as a set. The current version of CONIT, however, is able only to pass on without alteration any Boolean operations that an individual system can perform. CONIT also provides a browsing command (show index [term]), which, for any data base, provides a display of terms alphabetically near to term in the data base index of terms.

An example of an actual USER/CONIT dialog is included at the end of the report, with explanatory comments added in the rectangular boxes. Messages originating from CONIT have a single vertical line beside them; messages that have been translated (at least in part) are indicated by a double vertical line; and messages without any markings have been passed from the originating retrieval system to user without a CONIT translation.

Other commands not discussed above nor included in the sample dialog have also been implemented and are discussed in the reference at the end of the paper.

From our work thus far we are satisfied that the virtual-system concept of networking heterogeneous bibliographic data bases is technologically feasible. Further study is required, however, to establish the cost/benefits factor, the degree of virtualness that is needed in order to provide satisfactory service and what, if anything, could be accomplished in networking if software/hardware changes were made to host systems, rather than leaving them unaltered, as is now being done.

Reference: Marcus, Richard S., and Reintjes, J. Francis, The Networking of Interactive Bibliographic Retrieval Systems, Report ESL-R-656, March, 1976.

SAMPLE OF A USER/CONIT DIALOG

conit
Welcome to CONIT. For help on how to use CONIT you may type 'help' followed by a carriage return; otherwise, you may now type any CONIT command.
USER::
pick suny
Attachment successful.

INDICATES SUCCESS IN ATTACHING TIP PORT

SUNY/MEDLINE:
Connection completed.

INDICATES PROPER INITIAL RESPONSE SEEN

SS 1 is number for your next search set.

suny is connected successfully
You are now speaking in CONIT

INDICATES LOGIN COMPLETED

(THESE MESSAGES, AND OTHERS, COULD BE IN BETTER, MORE VIRTUAL FORMAT AS DISCUSSED IN REPORT)

USER::
find radiation

SUNY/MEDLINE:
Your search resulted in set1 which contains this many documents: (2684

SS 2 is number for your next search set.

USER::
show title docs1-3

SUNY/MEDLINE:

1
TI - :STATISTICAL EVALUATION OF LIGHT PROTECTION FACTORS:

2
TI - PROGNOSIS AND POST-THERAPEUTIC FOLLOW-UP OF BREAST CANCERS BY THERMOGRAPHY.

3
TI - MICROWAVE THERMOGRAPHY: A METHOD OF DETECTING SUBSURFACE THERMAL PATTERNS.

SS 2 is number for your next search set.

USER::
pick lms

LOCKHEED DIALOG SYSTEM PICKED

sent

INDICATES LOGIN SENT

HOST IS ONLINE
\$\$\$\$\$\$\$\$ RECONNECT @ 15:53:05

NEW FILE - DISSERTATION ABS. - SEE ?NEWS

USER::
pick data eric

FIIF1

SAMPLE OF A USER/CONIT DIALOG - continued

File reset: ERIC FULL-TEXT ED & EJ DEC-7

USER::
find radiation

Your search resulted in set2 454 RADIATION (PROCESS OF ENERGY

@ CANCELS LINE TO THAT POINT

USER::
show docs@show title docs1-3

NOTE SPECIAL AND IMPORTANT MESSAGE IN LOGIN

Msg from 1:DIALOG GOING DOWN FOR THE DAY
IN 5 MINUTES
1
EJ121600
Solar Heated Homes: They're Here

2
EJ121508
A Course in Nuclear Radiation for All High School Students

3
EJ121506
Energy Alternatives

USER::
pick sdc
sent
;
YOU ARE ON LINE L8A

SDC ORBIT SYSTEM PICKED

HELLO FROM SDC/ORBIT.
YOU ARE NOW CONNECTED TO THE ORBIT DATABASE.

SDC/ORBIT:
****
TODAY ONLY: GEOREF AND APIPAT WILL NOT BE AVAILABLE. SORRY
FOR THE INCONVENIENCE.
****

NOTE SPECIAL MESSAGE

SS 1 is the number of your next SDC/ORBIT search set.

USER::
pick data ntis

SDC/ORBIT:
-.
THE TIME IS NOW 6 :56 P.M. (EST), 01/16/76
YOU ARE NOW CONNECTED TO THE NTIS DATABASE.

SS 1 is the number of your next SDC/ORBIT search set.

USER::
find radiation

SAMPLE OF A USER/CONIT DIALOG - continued

```
SDC/ORBIT:
SS 1 is resulting set containing this many documents:(21223)

SS 2 is the number of your next SDG/ORBIT search set.

USER::
show title docs1-3


SDC/ORBIT:

TI  - Results of the Lyman alpha Measurements of the  Satellite Dial
      Ergebnisse der Lyman-alpha  Messungen des Satelliten Dial

TI  - Materials  Data Retrieval at Estec

TI   - Angular Distributions of Electrons of Energy Esub E Greater Than
      0.06 Mev in the J ovian Magnetosphere

SS 2 is the number of your next SDC/ORBIT search set.

USER::
pick nlm
sent
;
 LOGON IN PROGRESS AT 18:59:30 ON JANUARY 16, 1976
 *** BACK72  AT NLM AND SUNY CONTAINS 1972 CITATIONS ONLY.***
 TOTAL ACTIVE TSO USERS: 06
READY


TSO LINE 07E
ELHILL3 IS NOT AVAILABLE AT THIS TIME.
READY
```

NOTE IMPORTANT MESSAGE

```
USER::
pick suny
Can't log off nlm
```

CONIT DOES NOT YET RECOGNIZE UNAVAILABILITY MESSAGE AND GETS CONFUSED. ANALYST GOES INTO TRANSPARENT MODE TO CARRY OUT TASK "MANUALLY" (WITHOUT TRANSLATION)

```
USER::
sp host
Transparent mode.  All of your input will now be sent directly to nlm without
interpretation,except the 'speak conit' command which will cause CONIT  to
resume interpreting your requests.

USER::
logoff


      LOGGED OFF TSO AT  19:03:37 ON JANUARY 16, 1976+

DROPPED BY HOST SYSTEM
PLEASE LOG IN:
USER::
suny4

PASSWORD:bcn
```

• • •

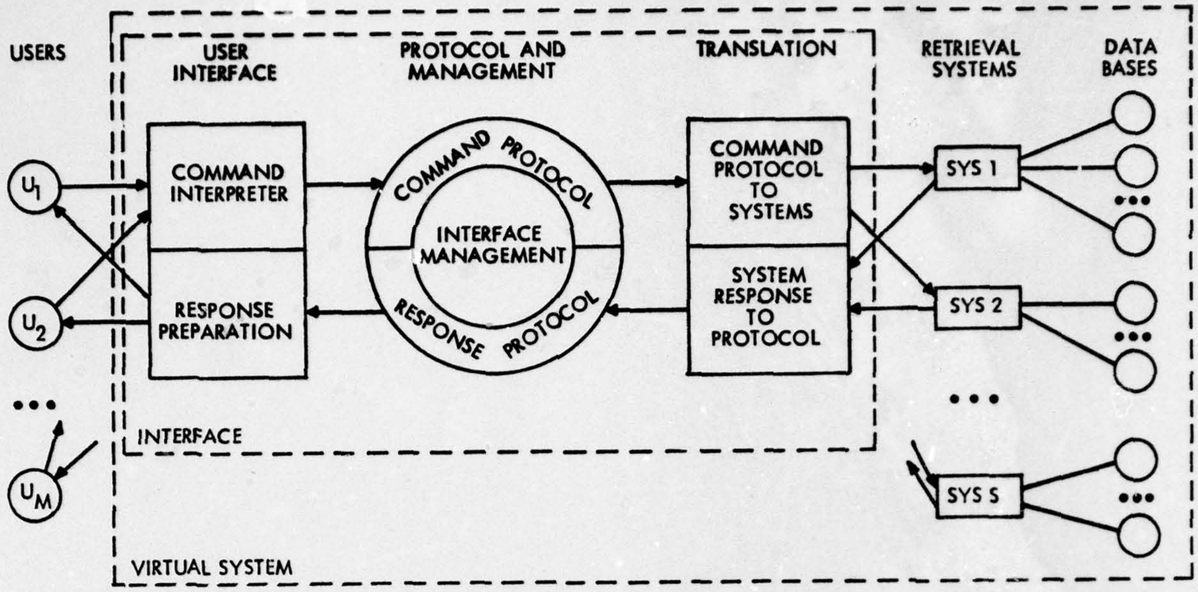


Fig.1 Logical diagram of virtual-system interface

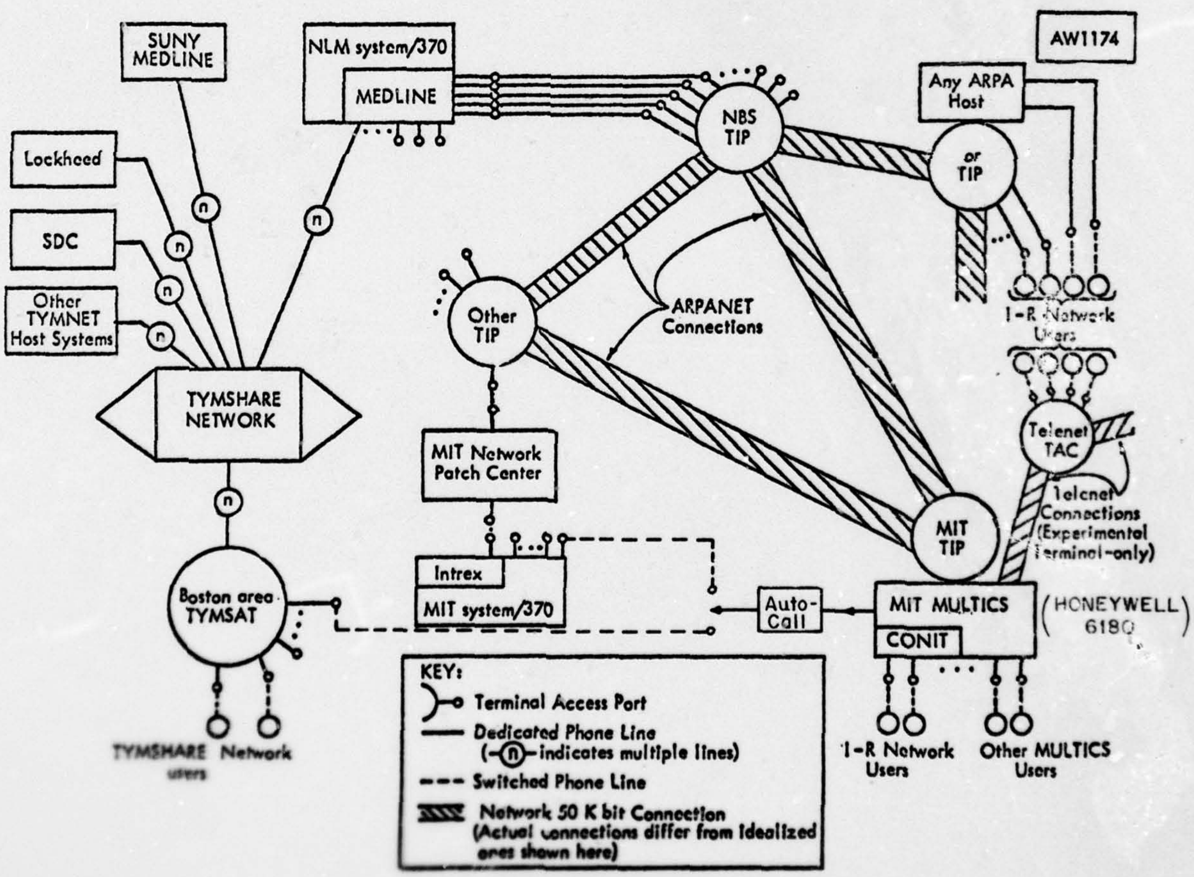


Fig.2 Computer interconnections for CONIT experimental interface

THE NATIONAL STANDARD REFERENCE DATA SYSTEM

Stephen A. Rossmassler
Office of Standard Reference Data
National Bureau of Standards
Washington, DC 20234

The National Standard Reference Data System is a coordinated, but decentralized, effort to increase the reliability and availability of numerical data used in and produced by the physical sciences and engineering. Individual data projects on specific technical subjects are established to extract, evaluate, and compile, in a systematic manner, all relevant data from the scientific journal and technical report literature. The evaluation process compresses the original data, and the systematic treatment aids the user in filling his data needs. Sophisticated data-handling capabilities -- including on-line information and data retrieval -- are developed in individual data centers and also in a central data systems design group.

Introduction

A major reason for the effectiveness of scientific research and development is the fact that today's scientists and engineers can build on what was done yesterday, and last year, and ten years ago. The results of past scientific work, carefully reported, appropriately published, are a basic part of the intellectual equipment of the technical worker. This is particularly true of the numerical results of scientific observation, the quantitative answers which Nature has given to our questions. We call these numbers "data". Data, combined with the technical understanding which comes from training and experience, are the tools of the scientist and engineer.

But a tool is useful only if it is reliable, and at hand to do the job. Data must be accurate and available if good results are to be achieved.

Such data are the concern and the reason for the existence of the National Standard Reference Data System (NSRDS). But why a special activity, since many scientific journals publish, abstract, and index a great volume of literature each year? What more is needed? From the user's viewpoint, much more is needed. The very bulk of the scientific literature poses great problems for the data-seeker. Even a carefully worded inquiry may result in more literature citations than he can read and assimilate. And when he has combed through fifty (or five hundred) papers and extracted the relevant data, how can he pick the best values? The published values for the thermal conductivity of copper, for example, range widely. Not all of the numbers can possibly be right. The answer is to involve the expert, and to encourage him to evaluate the data in his own field of special competence. Such a program provides reliable data. At the same time, because inaccurate or poorly documented data are weeded out, and inconsequential papers are discarded, data evaluation compresses the literature, often by a factor of more than 100, sometimes by a factor of 1000. In this way the data are also made much more accessible.

The NSRDS, as a technical information system, offers a number of retrieval concepts which warrant your attention today. The first feature is the one I have just mentioned -- the process of systematic thorough screening and evaluation of the original literature does much to facilitate data retrieval by the user. At the same time, each data compilation desirably contains citations to the original sources. Thus, for the scholar, both the expert and the non-expert, data compilations are an excellent point of entry into the literature on each special subject covered.

The second feature of interest is the information-handling capabilities which the individual specialized data centers develop in order to do their jobs efficiently.

The third feature involves the general-purpose data-manipulation procedures developed within the NSRDS for use by data centers and those who use their products. These two latter features will be described in more detail later in this presentation.

History and Mission

Numerical data compilation and evaluation, as a valid part of science, have a long history, full of notable contributions. The Periodic Table of Mendeleev, familiar to freshman chemistry and physics students, is a good example. Other early contributions include the Landolt-Bornstein Tables (which are still being updated and published) and many others. The National Bureau of Standards has taken part in this work for over half a century. The editor of the International Critical Tables, Dr. Edward Washburn, was a member of the NBS staff.

While such data evaluation projects were highly valuable, they lacked a central focus. There was no responsibility for planning, or for setting of priorities. To remedy these deficiencies, the Federal Council for Science and Technology, in 1963, enunciated a Federal Policy that there be a "National Standard Reference Data System." The National Bureau of Standards was given responsibility for the administration of the system, and for the coordination of the data programs of other Federal agencies. A small program management office, the Office of Standard Reference Data (OSRD), was established for this purpose.

This policy was augmented, in 1968, by the Standard Reference Data Act, Public Law 90-396, which expressed Congressional endorsement of the general undertaking. The Act also authorized the Secretary of Commerce to secure copyright on standard reference data compilations, and to sell them as part of a cost-recovery program. The copyright authority is not unique among Federal agencies, but it is unusual.

The general mission of the NSRDS is defined in these two documents -- the 1963 Policy, and the 1968 legislation. Refinements and added details have emerged with twelve years of operation. Basically there are five aspects to this mission, as follows:

1. To plan, fund, and administer a program of data evaluation in the physical sciences;
2. To publish compilations of data;
3. To coordinate data programs nationally;
4. To provide a focus for interaction with similar programs of other countries, and with international efforts; and
5. To improve the quality of laboratory measurements through provision of benchmark data, standards, criteria, and the like.

Operational Features

The data evaluation program is decentralized. In order to involve the participation of leading experts, wherever they work, the actual compilation and evaluation takes place in many widely separated groups. The fact that these groups vary greatly in size, organizational structure, and source of financial support leads to a certain degree of diffuseness in the system. However, it is instructive to recognize two broad types of components. The first is the continuing data center, which has a charter (formal or informal) to cover a certain technical area on a continuing basis. Ideally, such centers have an assurance of stable long-term financial support. The task of a data center is to search the world literature on a regular basis, to retrieve and index papers within its scope of interest, to extract the numerical data, and to carry out critical evaluation leading to publication of tables or reviews. Such centers are normally located in the environment of a scientific laboratory. Many of the centers are highly mechanized and have large, computer-based files from which bibliographic citations and data can be retrieved. Generally, they are able to respond to requests for specific information from the scientific public.

The other type of component is the individual scientist (or small group of collaborators) who produces a "one-shot" compilation or critical review as a part of what he regards as his normal scientific activity. Many valuable data compilations have been produced in this way. Such individuals do not consider themselves part of a formal data center, and there is generally no commitment for continuity or updating. The rapid growth of the scientific literature makes it increasingly difficult for an individual to do this type of compilation. However, the continuing data centers can serve a useful function by providing bibliographic back-up for individual scientists in other locations who wish to write critical reviews or do critical compilations of limited scope.

In the pattern which has emerged, it is clear that both types of components are essential for successful operation.

One of the primary goals of the Office of Standard Reference Data is to establish continuing data centers in all technical areas which fall within the scope of the program. Since the resources of OSRD are limited, every effort is made to encourage other Federal agencies and private organizations to participate in the support of these activities.

In the period since the establishment of NSRDS, data evaluation activities in other parts of the world have increased considerably. Formal governmental programs similar to NSRDS have been established in several countries. In the U. K., the Science Research Council administers a program which includes the support of a number of data centers in the physical sciences. In the Soviet Union, the Academy of Sciences supports several data evaluation projects, and the State Service for Standard Reference Data (GSSSD) has responsibilities similar to those of NSRDS for scientific and technical data. The Office of Standard Reference Data maintains liaison with these groups with the aim of avoiding duplication in data compilation projects and promoting maximum compatibility of output. To effect this liaison, proposals requesting support of projects in subject areas of mutual concern are exchanged for comment.

The establishment of the Committee on Data for Science and Technology (CODATA) in 1966 has provided a formal framework for international cooperation. CODATA is a committee of the International Council of Scientific Unions (ICSU). The Secretariat is located in Paris. The main purpose of CODATA is to encourage, on a world-wide basis, the production and distribution of critically evaluated numerical data.

Since the entire thrust of the NSRDS is user-oriented, much attention is given to interactions with users. Such interactions fall into three general categories:

1. Data products and services. These are discussed in detail below.
2. Feedback mechanisms, including surveys, interviews, questionnaires, symposia, and correspondence.
3. Advisory groups representing special user groups, specific scientific discipline interests, etc.

Products and Services

Most of the output of the NSRDS consists of compilations of evaluated data and critical reviews of the state of quantitative knowledge, each covering a specific (usually quite narrow) subject. These products are published through a variety of channels, including the following:

Journal of Physical and Chemical Reference Data. A quarterly journal containing data compilations and critical data reviews, published for the National Bureau of Standards by the American Institute of Physics and the American Chemical Society.

NSRDS-NBS Series. A publication series distributed by the Superintendent of Documents, U.S. Government Printing Office.

Appropriate publications of technical societies and commercial publishers. Handbook publications are a particularly important element of this category.

A recapitulation of the output of ten years of program operation showed 160 compilations, covering 36,000 pages and over 30,000 materials.

Recently the Office of Standard Reference Data has begun to disseminate some large data files in the form of magnetic tapes. These are sold through the National Technical Information Service. At present, three data tapes are available, on mass spectral data, atomic spectra, and crystallographic data.

Going beyond publications, an important category of information and data services is the answering of inquiries. Both the OSRD and its affiliated data centers (about 20 in number) attempt to answer, within restrictions imposed by limited staff and time, inquiries concerning data in the physical sciences and engineering. Information and data are provided from NSRDS publications and other available sources. At times references are provided; if available and appropriate, copies of publications or excerpts therefrom containing the requested data are furnished. OSRD normally receives about 1000 requests and inquiries per year. Under a recently established cost-recovery program, custom services which require a substantial amount of time are billed to the requestor.

Information Retrieval Technology

It has already been pointed out that the process of data compilation and evaluation improves the accessibility of data, as compared with their status in the original journal and report literature. The improvement takes place in several ways. First, the primary publication usually focuses on a research or development project of which data-gathering is only a part. Data results are frequently buried in the text, and are not identified in the abstract. Systematic compilation leads to greater visibility, and republication takes place through a medium explicitly devoted to data.

Second, data tabulation involves eliminating many of the non-data aspects of the original paper. The effect is a major reduction in volume, usually more than 100-fold. For example, in a recent issue of the Journal of Physical and Chemical Reference Data, an article on "Critical Region Parameters" cited and extracted data from 162 references; another on "Diffusion in Copper and Copper Alloys" derived data from 484 references. Such compaction is very helpful to the data-seeking user in his retrieval.

Third, the user gains greatly because the data compiler makes a commitment to cover the literature comprehensively. The latter's literature search is extensive and systematic. While no compiler claims perfection in coverage of sources, the compilation user is afforded the benefits of a search in which an acknowledged expert tried to do the job completely. Many data centers feel that they probably capture close to 90% of the pertinent literature.

Beyond this inherent improvement in retrieval capability, the NSRDS works both in its data centers and centrally to achieve improvements in information retrieval technology.

Data centers have a strong selfish interest in such improvements. They have a heavy work load, a responsibility to search both the past and current publications within their subject scope, and the difficult scientific task of data evaluation. Accordingly, they develop advanced and highly effective, though often very specialized, information retrieval systems. Typical features of such systems are the following:

1. Reabstracting of the original literature, with extraction or annotation and filing of data content.
2. Systematic indexing of data contents for computer searching.
3. Magnetic card or punch paper tape recording of data files, to reduce retyping.
4. Fixed-field formats for in-house records to assist in retrieval.
5. Automated preparation of bibliographies as a spin-off of in-house records.
6. Final publication via computerized typesetting from in-house files.

More broadly applicable are the data handling, manipulation, and retrieval capabilities developed by the OSRD Data Systems Design Group. The capabilities have recently been combined into a general-purpose data management system called Omnidata. This system has provision for: information and data retrieval; numerical, statistical, and graphical analysis; file definition and updating; encoding and decoding of data fields; multi-key sorting; arithmetic operations and general file manipulation. Going beyond the normally available routines for searching, arithmetic operation, statistical analysis, and report generation, Omnidata offers 36 modules, including such user-oriented procedures as Browse, Rename, Graph, and Tally, as well as file-building procedures, such as Check Sum, Screen, and Update.

A detailed description of this system will be available in a forthcoming NBS Handbook, "Omnidata: A Computer Program for Data Retrieval, Statistical and Graphical Analysis, and Data Management," by Hilsenrath and Breen.

Those who may want a more comprehensive overview of the National Standard Reference Data System are referred to a recent version of the status report "Critical Evaluation of Data in the Physical Sciences," NBS Technical Note 881. The report contains a list of data projects and data centers, a publication list, and other details. It is available from the Office of Standard Reference Data, or from the Government Printing Office as SD Catalog No. C 13.46.881. The price is $1.10, plus 25% postage outside the United States.

To summarize, the National Standard Reference Data System is a user-oriented information system based on data evaluation. The philosophy and operations of NSRDS demonstrate a number of information retrieval techniques which increase both availability and reliability of the data.

Synthesis and Distribution of Environmental Satellite Data

James I. Vette, National Space Science Data Center
Code 601, Goddard Space Flight Center
Greenbelt, Maryland, 20771

Summary

The activities of two National Oceanographic and Atmospheric Administration and two National Aeronautics and Space Administration facilities involved in the synthesis or distribution of space environmental data are reviewed. The data products, user services, and pertinent publications are given. The computer systems that support three of these facilities are discussed and some details of the synthesis procedures are given. Data from the following satellite series are included in the discussion: TIROS, NIMBUS, ESSA, NOAA, SMS/GOES, ATS, SOLRAD. Orbital position data for HAWKEYE 1, IMP H&J, VELA 5B, PIONEERS 6-9, PIONEER 11, HELIOS 1 & 2, and SOLRAD 11A & 11B in several coordinate systems are discussed.

I Introduction

The use of earth-orbiting spacecraft to obtain environmental data on a nearly global basis has increased tremendously in the past decade. In this same period, the growing sophistication of flight instruments and increased data rates require advanced techniques to synthesize and distribute effective data products to the various users. To detail all the spacecraft programs acquiring and processing environmental data is beyond the scope of this paper. However, to indicate some of the techniques that are being employed and to provide an idea of the types of environmental data available, specific examples have been chosen for discussion. These encompass meteorological, earth resources, near-earth charged particle, solar electromagnetic, and orbital position data. The techniques employed involve image processing, digital data processing and compression, photographic processing, interactive computer graphics, and computer generated photographs. The distribution ranges from on-line systems with real time data bases to synthesized products which are available upon request by mail or telephone.

The activity of four separate facilities will be presented. The Satellite Data Services Branch (SDSB) of the Environmental Data Services (EDS) National Climatic Center (NCC) of the National Oceanographic and Atmospheric Administration (NOAA) provides environmental and earth resources satellite data to users once the weather forecasting services have been provided. Emphasis will be placed on meteorological data from the NOAA operational series (known as ITOS before launch) and the geostationary spacecraft series SMS/GOES. The Space Environment Laboratory (SEL) Data Acquisition and Display System (DADS) is a real-time solar-terrestrial environment data system operated by SEL of NOAA's Environmental Research Laboratories. Data on solar x-rays, charged particles, and vector magnetic fields are obtained from SMS-1, SMS-2, and GOES-1 satellites in real time. Other spacecraft and ground-based measurements are available through the system. Two National Aeronautics and Space Administration (NASA) facilities, the Atmospheric and Oceanographic Information Processing System (AOIPS) and the International Magnetospheric Study (IMS) Satellite Situation Center (SSC), are both located at the Goddard Space Flight Center (GSFC). The AOIPS is used to process LANDSAT data for land use, hydrological, oceanographic, and geological studies and SMS/GOES data for meteorological research. The IMS/SSC processes both predicted and achieved orbital position data for a number of satellites in a variety of coordinate systems and determines interesting confluences of this ensemble of spacecraft.

The thrust of this paper is to provide some insight into the work being accomplished at these facilities and to indicate the distribution of the resulting data products. It is not surprising that computers play a vital role in these activities; however, it may be unexpected that three of these facilities rely heavily on mini-computers.

II The Satellite Data Services Branch

This facility is collocated with the operations center of NOAA's National Environmental Satellite Service (NESS) in the World Weather Building, 5200 Auth Road, Washington, D.C., 20233, and serves as a data source for (a) environmental data from TIROS, NIMBUS, ATS (I & III only), ESSA, NOAA, and SMS/GOES spacecraft, (b) earth resources data from LANDSAT spacecraft, and (c) photographs of the earth from SKYLAB missions. A more detailed summary of their holdings and services has been published recently (Reference 1).

The visible light and infrared data taken by a series of improving experiments on the spacecraft listed in (a) have been processed to produce photographs. In addition, the digital data from many of the later experiments are available on magnetic tape since each photographic step tends to degrade the data and many users wish to employ special digital processing. Currently SDSB receives daily about 250 negatives from the polar-orbiting NOAA spacecraft and about 250 negatives of full-disc and sector images from the SMS/GOES spacecraft. The processing of the original data are done by NESS and the final products are provided to SDSB for distribution.

The geostationary spacecraft provide full-disc and sector images from the Visible and Infrared Spin-Scan Radiometer (VISSR). The visible full-disc photographs are produced with a two mile resolution and the sector images with a one mile resolution. These products are important for studying global weather systems and provide the evolution and motion of cloud patterns. An example of the full-disc image is shown in Fig. 1. Computer generated hemispherical composites from the lower altitude NOAA satellites are produced

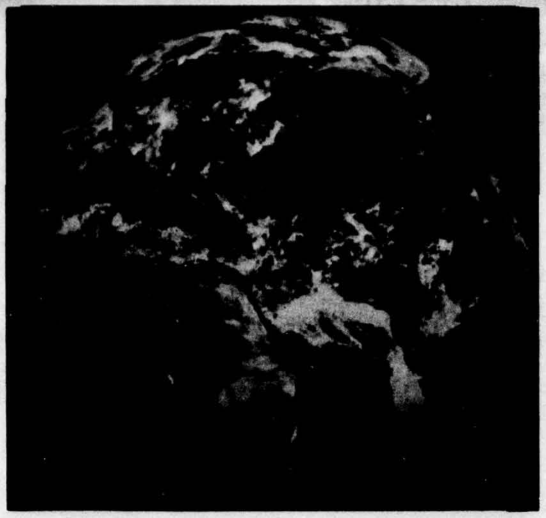


Fig. 1. Full-disc visible photograph of the earth. This picture was taken by the VISSR on SMS-1 on July 17, 1975 at 1700 UT with the visible channel (0.55-0.75 μm). The sub-satellite point was 1.5°S, 75.0°W. The instrument also has an infrared channel at 10.5-12.5 μm. The gridding and landmarking were done by computer during the picture construction at NESS. The instrument is capable of 0.9 Km resolution at the nadir but this product has about 3 Km resolution. Approximately 18 minutes scanning time are required for the instrument to obtain the data to produce this picture.

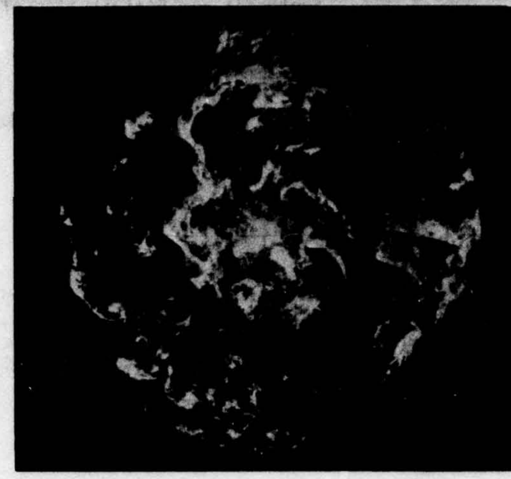


Fig. 2. Computer-generated Northern Hemisphere mosaic photograph. This composite was generated from the NOAA-2 Scanning Radiometer visible channel (0.52-0.73 μm) data acquired on 12 orbits during August 25-26, 1974. The GMT times mark the equatorial crossing grid so that nominal time for each swath can be estimated. The satellite is in a sun-synchronous orbit at an altitude of 1450 Km with an inclination of 101.7°. Using the Greenwich meridian one can determine the local time for each daytime swath is approximately 9-9:30 a.m. Resolution is 4 Km at the nadir and about 10 Km at 1600 Km to either side of the nadir.

in polar stereographic projections gridded with circles of latitude and rays of longitude; an example is shown in Fig. 2. Mercator-projection composites for middle and low altitudes are also generated with a rectangular latitude-longitude grid (Fig. 3). Infrared data are used to generate both nighttime and daytime composites, while visible data are only useful for daytime. The composites are produced by the Scanning Radiometer (SR) on the NOAA satellites; additional data are obtained from the Very High Resolution Radiometer on board NOAA 3 and 4, which has 1 Km resolution at the nadir for both visible and infrared channels. A detailed description of the procedures used to produce the digitized mosaics is available (Reference 2).

The data formats for the various products include photoprints (usually 10" x 10" on matte paper), negatives, transparencies, positive or negative 35 mm reels, slides, and movie loops of changing weather patterns. A Catalogue of Operational Satellite Products has been published (Reference 3). EDS publishes

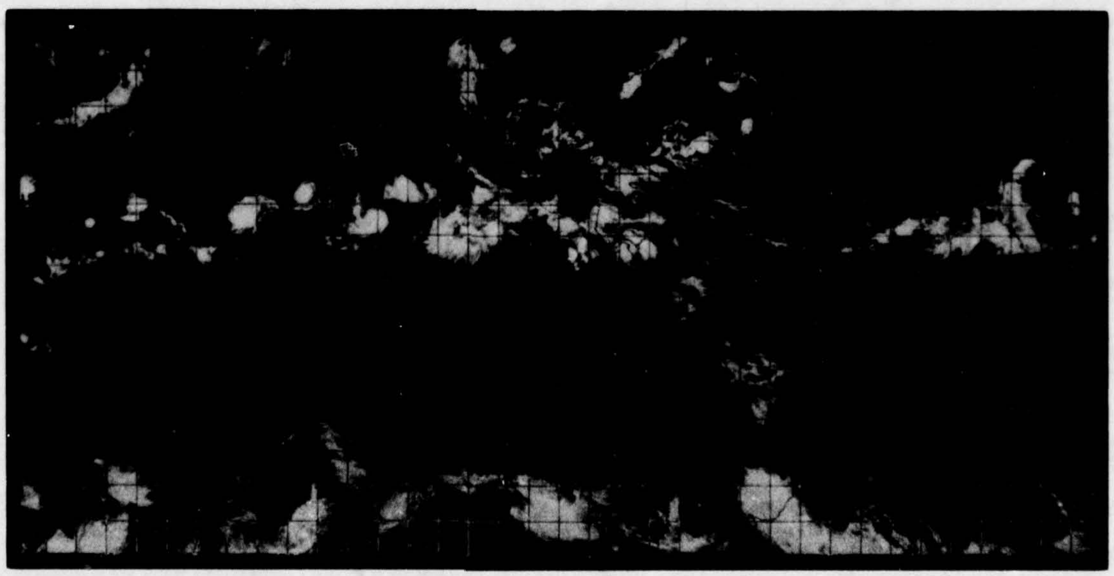


Fig. 3. Mercator-projection mosaic photograph. This composite was generated from the NOAA-2 Scanning Radiometer infrared channel (10.5-12.5 μm) data acquired during August 25, 1974. The numerous storm systems at low northern latitudes are clearly seen. Resolution is half as great as for the visible channel shown in Fig. 2.

"Key to Meteorological Records Documentation No. 5.4 (Environmental Satellite Imagery)" as a monthly issue describing data available from NOAA satellites which includes 5" diameter daily mosaics (visible and infrared) for both Northern and Southern Hemispheres.

III SELDADS

SELDADS is a real-time solar-terrestrial environment data system operated by SEL located at 325 South Broadway, Boulder, Colorado 80302. A detailed description of the system has been given recently by Williams (Reference 4). The general concept is displayed in Fig. 4. A variety of experiments, both ground-based and satellite, furnish data to the system.

The satellite data provide measurements of solar x-rays, solar wind plasma, electrons, protons, alpha particles, and vector magnetic fields at synchronous altitudes. More details of these experiments are presented in Table 1. The data from SMS/GOES are collected directly in real time while the data from the other spacecraft are sent through a communication system made up of the USAF Astrophysical Telecommunications Network (ATN), a USN commercial line, other dedicated commercial lines, a dedicated meteorological line to Moscow, and the International Ursigram World Days Service (IUWDS) network. The SMS/GOES data are converted immediately, processed, recorded on magnetic tape, and inserted into the real time data system. The data base consists of one minute averages for the most recent four days and five minute averages for 32 days. Data purged from the data base are placed on magnetic tape and archived at the National Geophysics and Space Data Center (NGSDC)/World Data Center (WDC)-A for Solar Terrestrial Physics in Boulder. Such tapes will also be available from the National Space Science Data Center/WDC-A for Rockets and Satellites at GSFC.

Near real-time data of polar cap solar proton fluxes are received from a Solar Proton Monitor on board NOAA spacecraft. These data arrive at SEL through the ATN and are inserted into the on-line data base. Fluxes observed across the polar caps are maintained for the most recent eight days, with polar cap averages being maintained for one month before being sent to the archive.

The data from SOLRAD 11A or 11B is processed by the Naval Research Laboratory and sent over the USN line. Priority is given to the satellite that is in the interplanetary medium; since these two satellites are maintained about 180° apart, one will always be outside the earth's bow shock. The final phasing of these two satellites was completed on July 20 but station keeping will be required periodically.

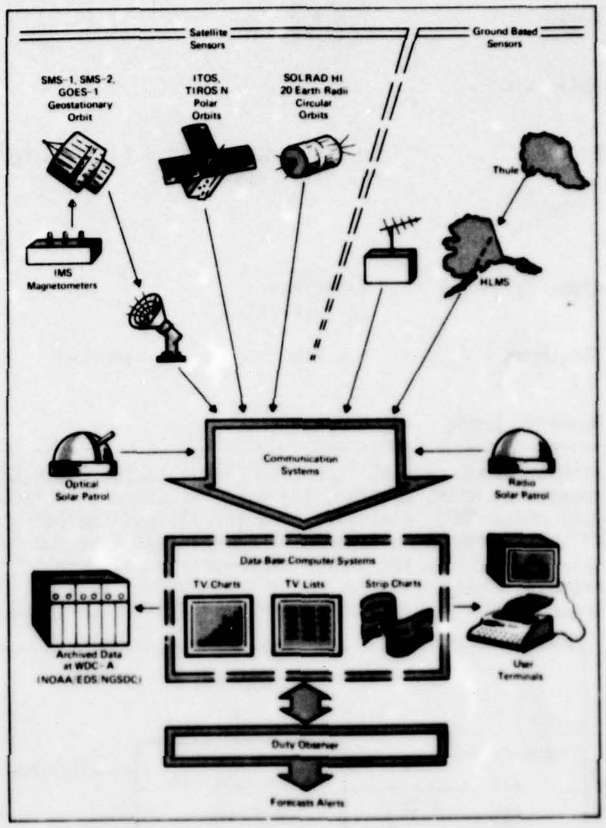


Fig. 4. Schematic overview of SELDADS. The geostationary orbiting spacecraft transmit data directly into the system. The NOAA (ITOS) solar proton data are sent over the ATN (see text), which is part of the communications system. SOLRAD 11A and B (SOLRAD HI) data are sent over a USN commercial line often processing at NRL. TIROS N is a follow-on series of satellites which will replace NOAA satellites; the first launch will be early in 1978. Various aspects of the system are discussed in the text.

There are a number of ground-based observations which are fed into SELDADS. In fact, over 75 stations around the world supply data of the following types: total electron content; hydrogen alpha (6563 Å) solar events, features, and patrol; discrete frequency solar radio flux and bursts; spectrographic solar radio events; solar calcium plage observations; coronal intensities; white-light sunspot observations; optical auroral observations; auroral radar backscatter; ionosonde observations; high-frequency radio path signal strengths; sudden ionospheric disturbances; high-latitude riometer data; and magnetic field observations. Much of the ground-based data come in through the world-wide IUWDS network.

Beginning in 1977 and extending through the period of the International Magnetospheric Study (IMS), which ends December 31, 1979, an extensive set of 25 magnetometer stations located in the Artic and across the Americas and the Pacific will relay data through the SMS/GOES satellites to SELDADS. These IMS magnetometer data files will be maintained at the original 10 sec resolution for four days and one minute resolution for 32 days before being dumped to magnetic tape and archived.

All incoming data are handled by the computer system shown in Fig. 5. The data base is contained on the two 58 megabyte disc drives. The two Nova 1200 mini-computers are connected by a multiprocessor communications adaptor (MCA) and can transfer data to and from each other at a 75 kilobyte rate. The new computer will have floating point arithmetic, two 10 megabyte disc drives, and a 70-character-per-line, 300-lines-per-minute printer. Display capabilities at the facility include interactive and call up CRT displays, strip charts, word messages, and printer output. Computer (1) in Fig. 4 has input/output interfaces for (a) acquiring data from two SMS/GOES satellites, (b) acquiring data from the USAF ATN line, (c) handling a 2400 baud link between SELDADS and the Global Weather Center at Offutt Air Force Base, (d) interfacing

Table 1. SELDADS Satellite Data Inputs

Satellites (Agency)	SMS/GOES (NASA/NOAA)	NOAA (ITOS) (NOAA)	SOLRAD 11 A/B (USN)	TIROS N (NASA/NOAA) 1978 Launch
Data Reception	2 satellites simultaneously	1 satellite of the series	1 of 2 satellites	2 satellites simultaneously
Orbit	Geostationary	1450 km, 101.7^o sun-synchronous	circular, 21 Earth radii	1450 km, 101.7^o sun-synchronous
X-rays	0.5-4 Å 1-8 Å	- - - -	3-150 keV (6 intervals) 0.5-1350Å (8 intervals)	- - - -
Solar Wind	- - - -	- - - -	Density, Velocity, Temperature	- - - -
Protons	0.8-500 MeV (7 intervals)	$\geq$ 10,30,60 MeV	20 keV-160 MeV (16 intervals)	0.3-2500 keV (6 intervals) $\geq$ 10,30,60 MeV 370 -> 805 MeV (4 intervals)
Alpha Particles	4-392 MeV (6 intervals)	- - - -	20-460 MeV (5 intervals)	640 -> 850 MeV (2 intervals)
Electrons	$\geq$ 2 MeV	$\geq$ 140 keV	11-1510 keV (11 intervals)	0.3-20 keV $\geq$ 30,100,140,300 keV
Magnetic Field	3 components	- - - -	- - - -	- - - -

with a communications line (DIG HIWAY) to the Space Environment Services Center (SESC), which is also located at NOAA, Boulder, (c) handling an on-site teletype, and (f) handling a remote user through a data phone (DP) terminal. Computer (2) has similar interfaces shown in Fig. 4 and sends data to a RAMTEK CRT device which is capable of generating eight low resolution (256 bit by 240 lines) and two high resolution (512 bit by 480 lines) alphanumeric and graphic displays. The final system will have three 9-track magnetic tape drives and one 7-track drive. The SEL Computer Terminal (CT) will allow access to a high-speed printer, additional terminals, and the Environmental Research Laboratories CDC-6600 Computation system.

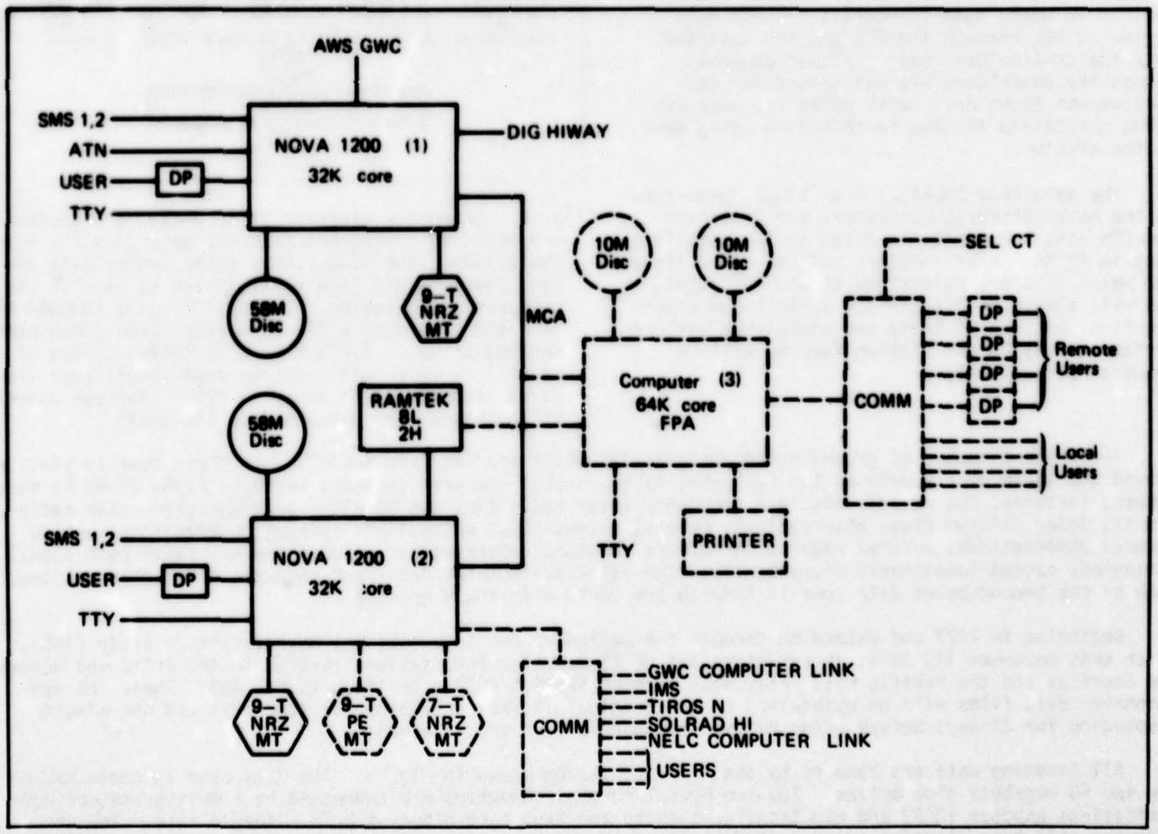


Fig. 5. SELDADS Computer Systems. Equipments which existed in 1975 are shown as solid line blocks; dotted line blocks represent equipments which are being procured and will be operational in the fall of 1976. The line labled DIG HIWAY is a communications link which connects the system to the SESC data displays.

The required equipment for a remote user is a 110 baud (10 characters/sec in ASCII Code) acoustically coupled terminal operating on a full duplex line and a coupler in original mode compatible with a Bell 103-C data set. Normal data dissemination procedures include direct computer-to-computer links with the real-time data system, user terminals, telephones, teletype reports and forecasts, a weekly summary of solar terrestrial environment activity issued by SESC, and special requests (Reference 5). The remote user is restricted to a print-out display on a real-time basis, unless he links the data into his own computer. For specific time periods users may request that CRT graphs be transmitted by telecopier over the telephone. Of course, this requires that the user have a compatible telecopy receiver compatible with SESC's Xerox 4001 equipment. All products of the facility are available through the mail.

Besides making this extensive and current data base available to a user, it is possible to obtain some synthesized products. Magnetograms from individual stations are useful in special cases but the stacked magnetograms shown in Fig. 6 are much more useful in understanding the development of geomagnetic phenomena. For a chain of stations lying along a geomagnetic meridian (three such chains will be operational during the IMS), the user will be able to call up the results of an electrojet current model calculation which is consistent with the observations. At one specific time the variation along the meridian can be displayed. Another mode provides the time history of the currents where the central latitude, width, and current intensity for each electrojet are given.

Much of the data available through SELDADS was not previously accessible until months after data acquisition. This rapid dissemination of space environmental data makes new research more effective and provides a rapid understanding of such things as radio blackouts, auroral displays, solar activity, and the terrestrial effects of large geomagnetic storms, where electrical power transmission can be impacted.

MAGNETOGRAMS OF MAGNETIC STORMS

JUNE 10-12, 1973

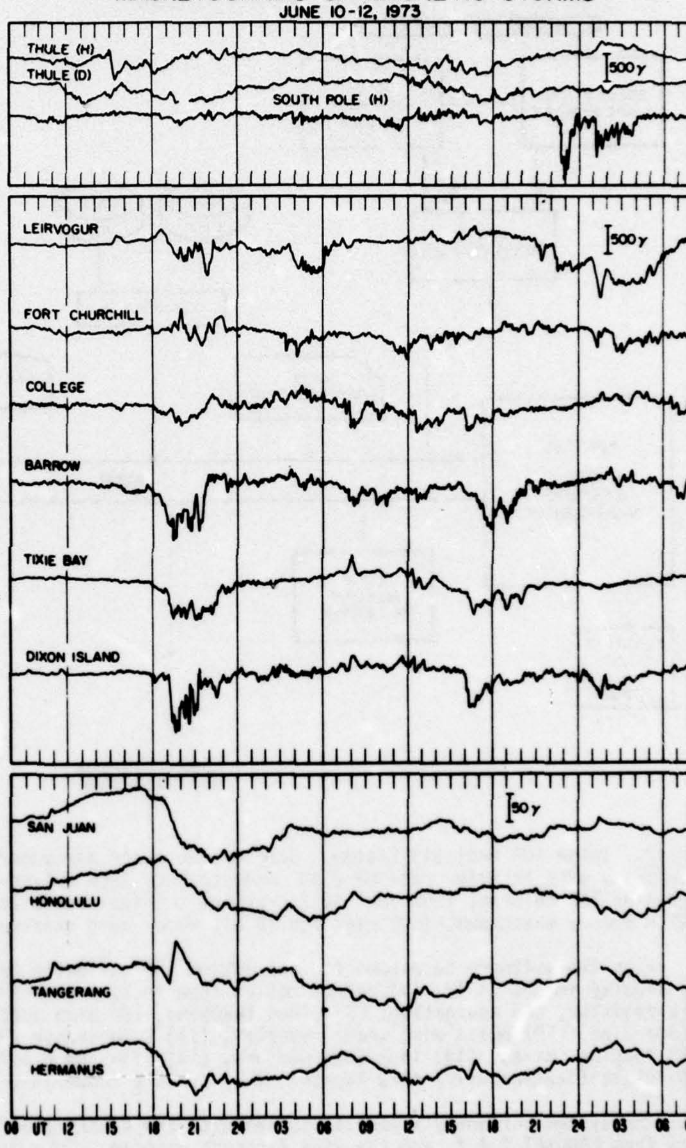


Fig. 6. Stacked magnetograms during magnetic storms. The upper panel contains data from stations very near the magnetic poles. The middle panel contains Northern Auroral Zone stations and low latitude magnetometer data is given in the bottom panel. This synthesized data product is useful in the study and evolution of magnetic storms and sub-storms.

IV AOIPS

AOIPS is an interactive system designed to provide image display, analysis, and data management for the NASA investigators in the disciplines of meteorology, hydrology, earth resources, and oceanography. This digital image processing system has been developed by the Computer Systems Branch of the Information Extraction Division, Applications Directorate, GSFC. The majority of satellite experiments involved in the above disciplines produce a series of line scans in one or more bands of the electromagnetic spectrum to provide a data format similar to that used to produce a TV picture and to that used in VISSR and SR instruments described in the previous section. The more precise systems such as AOIPS handle the data in a digital fashion rather than analog. The resolution of the resulting image depends on the number of lines and the number of pixels in each line as well as the field of view of the instrument. The other parameter is the number of levels of intensity (shades of gray) per pixel. The total number of bits generated by such a satellite instrument can be quite large and efficient processing is extremely important. AOIPS provides for image information extraction of 512 lines x 512 pixels.

The important functions which such a processing system must provide are the following. Registration is important in being able to superimpose the images obtained with multispectral devices so that a composite color or false color photograph can be obtained. There is obviously noise or interference which intermittently produces errors in certain pixels and it is desirable to correct these sporadic effects. The ability to overlay a grid or landmark information is necessary to carry out a proper analysis of certain data and to display the results in a meaningful way. In the study of cloud patterns, ice flows, land use, soil moisture

and other similar topics the ability to present a time lapse is extremely useful. This requires data management techniques to access data which provide pictures of the same scene at different times and to put on image display devices the appropriate elements to understand the general time evolution of the desired quantities. Classification of the ensemble of bits that comprise a digital picture is useful in order to reduce the number of choices that one needs to retrieve and study desired images. Finally, display of the resultant selection and manipulation of the elements are important. This involves producing outputs on printers, plotters, TV displays, CRT's, image recorders, and photographic film.

The hardware configuration that has been chosen to carry out these extensive tasks is presented to show elements of a typical system. A more detailed description of this system has been given by Dalton (Reference 6). A modified General Electric Image 100 Interactive Multispectral Image Analysis System with provisions for a TV scanner which digitizes hard copy images, maps, overlays, and other information is shown in Fig. 7. A more extensive data management and display system which can link into the Image 100 system and also link into the large IBM 360/91 GSFC computer through a 4800 baud half-duplex bi-synchronous line is blocked out in Fig. 8. The Digital Equipment Corporation (DEC) PDP-11/45 mini-computer controls the Image 100 Analyzer Console and single color graphics output can be displayed on Textronix 4012 terminals. The PDP-11/70 mini-computer controls the GSFC designed image display terminal. Important parameters of the equipment are given in the figure captions. A Dicomed recorder, which is not shown, is used to produce both black and white and color photographs of synthesized products.

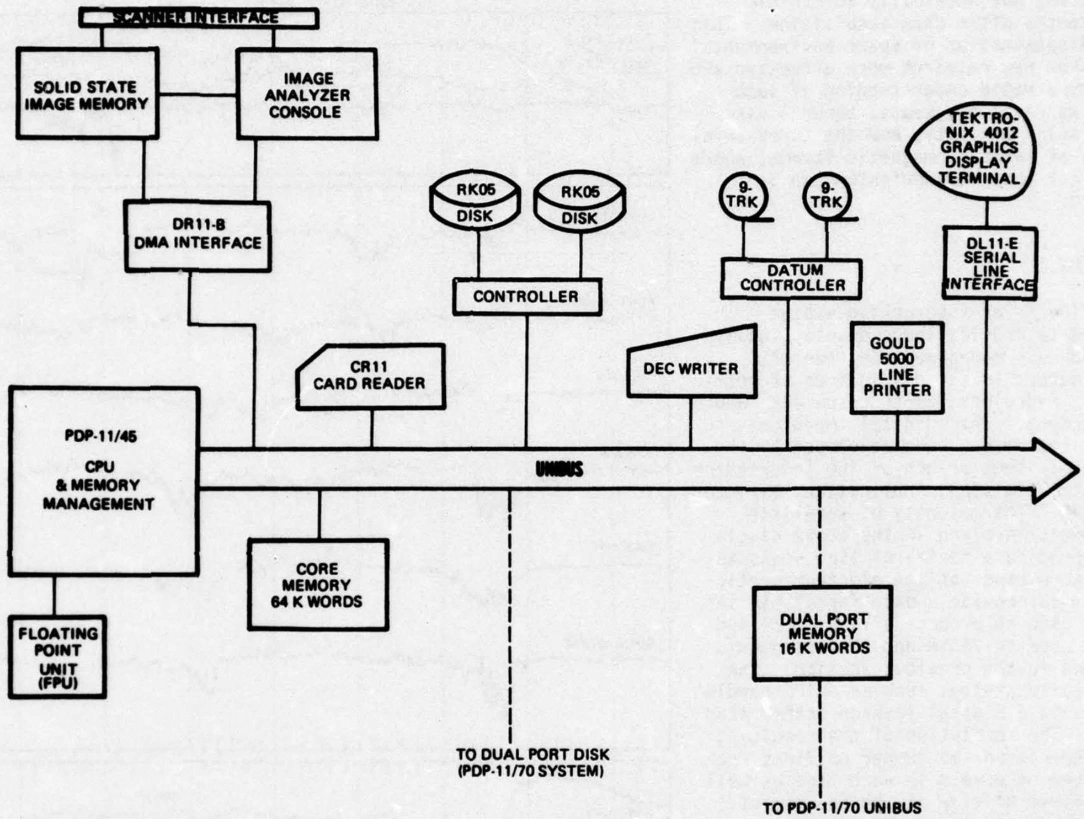


Fig. 7. Image 100 Analysis System. Some of the basic parameters of the elements are: Disk-2.4 megabyte capacity with transfer rate of 0.18 megabyte/sec; tape drives-9-track, 125 ips, 800/1600 bpi; line printer 132 columns, 1200 lpm; 4012 graphics terminal operated at 1200 baud; basic cycle time of PDP 11/45 with memory management-1.07 μsec for 16 bit word; card reader-300 cards/min.

With the software developed for the system the following capabilities exist: (1) list image files, (2) display images on TV, (3) reduce master tone to TV size, (4) display subarea (zoom), (5) perform landmark register, (6) navigation, (7) cloud tracking, (8) wind output, (9) image registration, correction, landmarking, (10) build wind vector overlay, (11) superimpose vector overlap on image, (12) flying reseaus, (13) looping images, (14) list wind vectors, (14) film image output, (15) produce Calcomp plot of output, (16) classification, (17) time lapsing, (18) feature computation/analysis.

The system currently supports studies with the VISSR data from SMS/GOES and the Multispectral Scanner data from LANDSAT 1 & 2, and the NASA Aircraft Program. Some experiments on Nimbus G and HCMM (both 1978 launches) will also be processed by AOIPS. These studies encompass forest inventories, dismal swamp signatures, power line routing, census urban area delineation, watershed survey, water resources management, soil moisture analysis, oceanographic thermal signatures, and severe storms.

One synthesized output from AOIPS is the divergence of the wind velocity field at low altitudes during Hurrican Elaine which is superimposed on the clouds. The divergence contours are shown in Fig. 9 in units of 10^{-6}/sec where H marks positive (inflow) peaks and L marks negative (outflow) peaks.

AOIPS has not been in existence long enough to have standard data products flowing out. However, as these products are developed they will be made available through data distribution groups such as NOAA's SDSB or NASA's NSSDC.

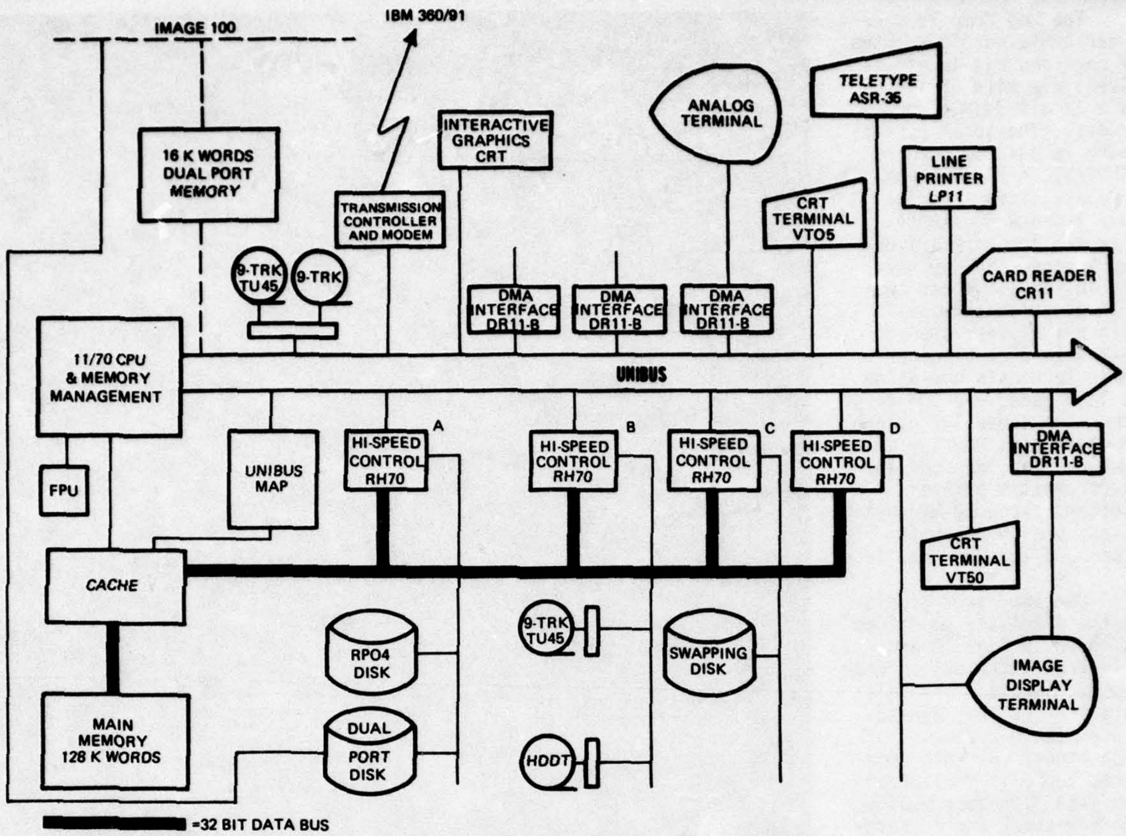


Fig. 8. GSFC-designed color image display and analysis system. Some of the basic parameters of the elements are: disks-88 megabyte capacity with transfer rate of 0.8 megabyte/sec; tape drives-9-track, 75 ips, 800/1600 bpi; line printer-132 columns, 300 lpm, card reader-300 cards/min; teletype-10 baud; VT05 alphanumeric-300 baud; VT50 alphanumeric-1200 baud; graphics CRT-1200 baud, image display terminal-2 megabyte/sec; basic cycle time of PDP 11/70 with memory management and interleaving-0.69 μsec for 32 bit double word; and High Density Digital Tape-14 channels, 7½ ips.

It is clear that exciting and useful products will be available in the near future.

V IMS/Satellite Situation Center

The Satellite Situation Center was established within NSSDC/WDC-A at GSFC to support the IMS, an international program involving some 44 countries in studying the earth's magnetosphere by means of balloons, sounding rockets, aircraft, ships, satellites, and ground-based measurements during the period 1976-1979. Although there are dedicated satellites being launched by several countries, including the core satellites GEOS (European Space Agency) and ISEE A/B & C (NASA), it is important to utilize all available in-situ measurements by older spacecraft when the confluence of positions provides rare or unique situations. The IMS program has been discussed recently in several articles (References 7, 8, 9, 10).

The main role of the SSC is to study the predicted orbits of some 50 or more spacecraft which have operable magnetospheric experiments and to identify interesting configurations of this ensemble well enough in advance so that the IMS scientific community can plan supporting measurements by balloons, rockets, etc. and can insure that the important data from the older satellites be acquired during the periods of interest. A more detailed description of the SSC and its activities is given in Reference 11.

The most useful satellite orbits to study for the above purpose are high altitude ones, where apogee altitudes are at least 12 earth radii (R_E). Fortunately most such satellites have perigees high enough that the atmospheric drag is negligible and orbital positions can be predicted accurately 6 to 12 months in advance. Based on the most recent orbit elements at the time, predicted orbit tapes giving geocentric coordinates as functions of time are generated for the desired intervals by the Operational Orbit Support Branch at GSFC and provided to the SSC. These data are then read onto a 25 megabyte disc to form the basic on-line data base for this work. Software has been developed to search this data base to determine which satellites are nearly simultaneously passing through regions of the magnetosphere which are of primary interest. These regions are the bow shock, magnetopause, magnetosheath, magnetotail cusp, interplanetary, and neutral sheet. Such regions can be defined mathematically, although in practice the boundaries move depending upon the solar wind and interplanetary magnetic field parameters. Several coordinate systems must be employed to carry out these queries and the geocentric coordinates of the satellites are converted to the appropriate system.

The results of the above computations must be displayed appropriately in the various coordinate systems. In order to do this efficiently, interactive graphics terminals (Textronix 4014) connected to a Mod Comp IV computer through a 9600 baud line are employed. The resultant graphs with appropriate annotation are photographed from a Textronix 611 slave CRT with an automated 35 mm camera.

The Mod Comp IV computer contains 512 K bytes of core and has an effective cycle time of 1.4 μsec for a 32 bit double word access. The total system, which is used to support NSSDC/WDC-A for Rockets and Satellites includes: four 9-track 800/1600 bpi, 125 ips tape drives; two 7-track, 556/800 bpi, 125 ips tape drives; two 4014 graphics terminals with 611 slaves; twelve Textronix 4023 alphanumeric terminals operating at 1200 baud; one 300 card/min reader; one 132 column, 96 character, 200 lpm impact printer; one 500 lpm electrostatic printer/plotter; three 50 megabyte discs; one 25 megabyte disc; and the Mod Comp IV.

Fig. 9. Divergence of wind velocity field during Hurricane Elaine. The contours are superimposed on low altitude clouds that were used to derive velocity field. Units of the divergence are 10^{-6}/sec. The data were taken over the period 1330-1400 UT on September 12, 1974. The ordinate is geographic latitude in degrees and the abscissa is east longitude in degrees.

One important display of the satellite positions is shown in Fig. 10 and illustrates the period when Imp H, Hawkeye 1 and Vela 5B all cross the magnetopause within 2.2 hours of each other. At this same time, Imp J is monitoring the interplanetary medium. The bow shock and magnetopause boundaries are drawn automatically. The time period and satellite are selected and the orbital arc is then drawn by the computer. The placement of the labels for each satellite arc, the boundaries, and the times associated with the tic marks are selected by the operator. The ecliptic latitudes are obtained from other computations but are indicated at the various times so that the three dimensional information on the magnetopause boundary crossing is displayed. Another view of this situation for day 104 of 1977 is shown in Fig. 11. As stated in the caption, more accurate magnetopause and bow shock crossing times can be determined from this presentation. Such crossing times are annotated in the legend at the top of the graph. The most concise presentation for this interval displaying all pertinent information to tne required accuracy is given in Fig. 12. It is easy to determine at a glance the region of space occupied by each of the four satellites as a function of time and the pertinent local time, latitude, and distance information to give a three dimensional view of the satellite situation.

Prior to the start of the IMS, the SSC studied the orbits of Imp H & J, Hawkeye 1, Vela 5B, 6A, & 6B during 1976 and identified 18 intervals ranging from 13 hours to 73 hours in length when these satellites were simultaneously (within 2 hours) traversing magnetospheric regions which should provide interesting measurements to further our understanding of the complex phenomena. These intervals were adopted as IMS Special Satellite Periods. A publication was distributed to all IMS participants showing the detailed positions of all the above spacecraft during these intervals (Reference 12). At some later date when the actual positions of the satellites during 1976 are determined accurately from tracking data, similar plots to those shown in Figs. 10-12 will be produced in addition to other types of plots which show the satellites relative to the neutral sheet and to the cusp region. In addition, some important new satellites, SOLRAD 11A & 11B and USSR's Prognoz 4 were launched in 1976 and will be included in the achieved orbit films.

For 1977 the SSC is currently studying positions of IMP H&J, Hawkeye 1, Vela 5B, SOLRAD 11A & 11B, the moon (scientific instruments left during Apollo landings are still operating) and ISEE A/B to determine the most interesting periods for data collection. When these studies are completed, reels of 35 mm film will be available to the IMS community on request and two hard copy reports (one for each 6 months) of the daily regionplots shown in Fig. 12 will be distributed.

It is through this synthesized positional data that more meaningful satellite environmental data can be obtained and made available to interested users.

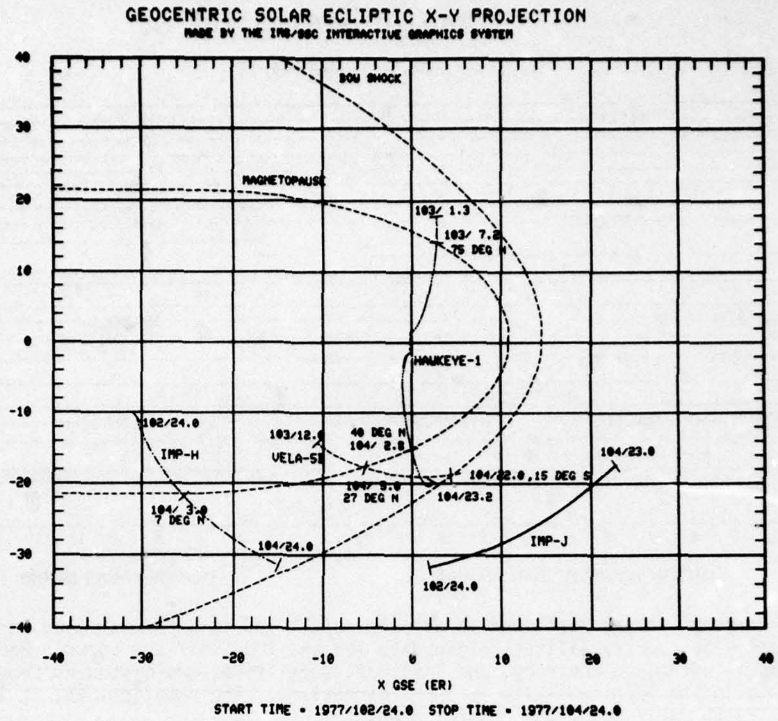


Fig. 10. Example of orbit projections for IMP-H, IMP-J, VELA 5B, and Hawkeye 1 by cylindrical rotation about the x-axis in the solar ecliptic coordinate system for days 103 and 104, 1977. A cursor placed at any point on the orbital arcs allows the user to produce a time tic and then the cursor is used to place the time of this point (day/hour UT) on the plot.

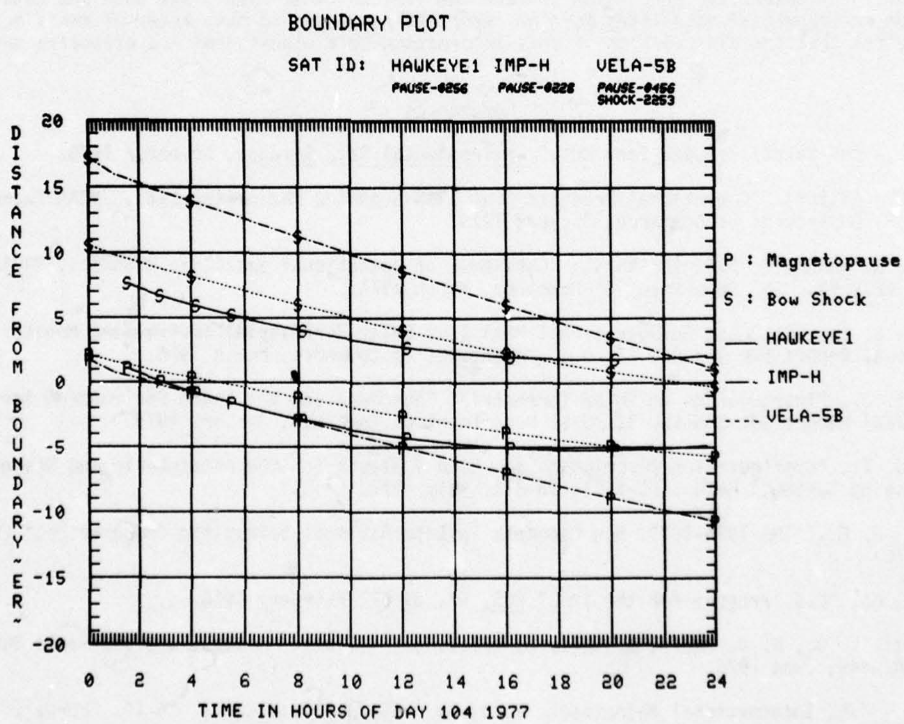


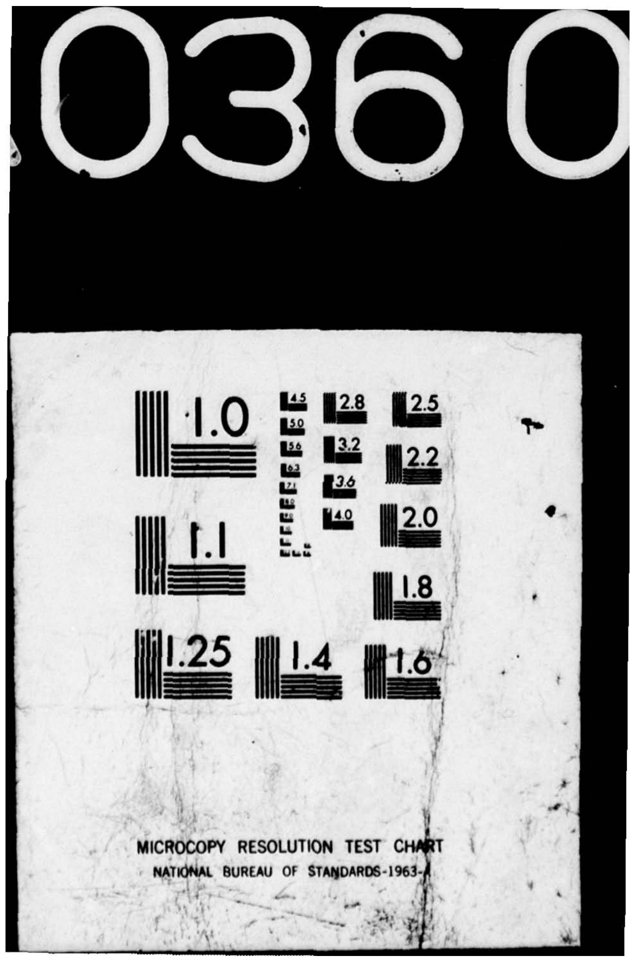
Fig. 11. Distances from the satellite (moving along the time axis) to the magnetopause and the bow shock as functions of UT. Crossing times can be determined more accurately from this type of plot than that in Figure 1; greater accuracy is not necessary since all boundaries move in response to changes in the interplanetary medium. Positive distances are on the concave side of the surface; negative distances are on the convex side.

AD-A036 030 ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT--ETC F/G 5/2
ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION --ETC(U)
DEC 76

UNCLASSIFIED AGARD-CP-207 NL

2 OF 2
AD
A036030

END
DATE
FILMED
3-77


0360
1.0
4.5
5.0
5.6
6.3
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.1
1.8
1.25
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SATELLITE REGIONS PLOT

DAYS 103 AND 104, 1977

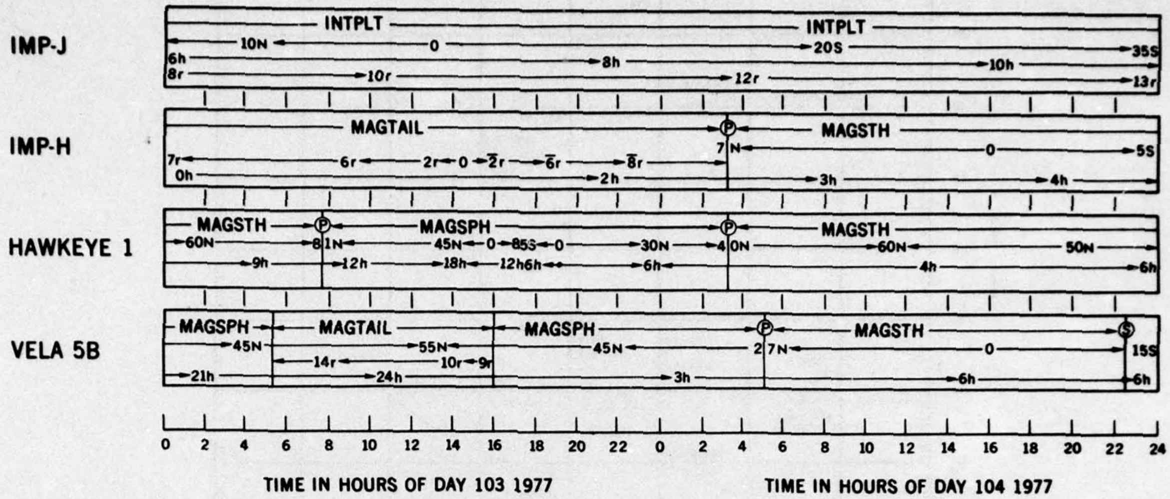


Fig. 12. The bar chart gives the positions of all the satellites in a concentrated format. In this chart the regions occupied by four satellites during days 103 and 104, 1977 are shown. Boundary crossings can be identified easily with good accuracy, and latitude, local time, and distances from neutral sheet or bow shock are given to locate each satellite in three dimensions. The notation, $\bar{2}$r, at 1600 UT on day 103 means IMP-H is 2 earth radii below the neutral sheet.

V Concluding Remarks

The purpose of this paper has been to convey techniques and equipment currently being utilized to handle, process, synthesize, and distribute the environmental data obtained from satellites. All of the facilities presented here except AOIPS are directly involved with data dissemination. Appropriate addresses have been provided so that any interested person can contact the various facilities. The data and data products resulting from environmental satellites form an important resources in many areas of man's activity. Therefore, it is vital that the distribution of this information be a wide-spread and effective operation.

References

1. Berry, L., "EDS Satellite Data Services," Environmental Data Service, November 1975.

2. Conlan, E. (Editor): "Operational Products from ITOS Scanning Radiometer Data," NOAA Technical Memorandum NESS 52. U.S. Department of Commerce, October 1973.

3. Hoppe, E. R. and A. L. Ruiz (Editors): "Catalogue of Operational Satellite Products, "NOAA Technical Memorandum NESS 53, U.S. Department of Commerce, March 1974.

4. Williams, D. J., "SELDADS; An Operational Real Time Solar-Terrestrial Environment Monitoring System," NOAA Technical Report ERL 357-SEL 37, U.S. Department of Commerce, March 1976.

5. Mangis, S. J., "Introduction to Solar-Terrestrial Phenomena and the Space Environment Services Center," NOAA Technical Report ERL 315-SEL 32, U.S. Department of Commerce, January 1975.

6. Dalton, J. T., "Configuration Description and Load Analysis for the Atmospheric and Oceanographic Information Processing System," NASA GSFC-X-933-75-217, July 1975.

7. Roederer, J. G., "IMS 1976-1979: New Concepts in International Scientific Cooperation," EOS, 57, 6-8, January 1976.

8. Manka, R. H., "U.S. Program for the IMS," EOS, 57, 63-67, February 1976.

9. Lanzerotti, L. J., R. D. Regan, M. Sugiura, and D. J. Williams, "Mangetometer Networks During the IMS" EOS, 57, 442-449, June 1976.

10. Knott, K., "The International Magnetospheric Study," ESA Bulletin No. 4, 16-19, February 1976.

11. Sugiura, M. and J. I. Vette, "The IMS Satellite Situation Center," to be published in EOS.

12. ____________, "Special IMS Periods for 1976" IMS/SSC Report No. 6, Decmeber 1975, NASA, GSFC.

DEVELOPMENT AND APPLICATIONS OF SPATIAL DATA RESOURCES IN ENERGY RELATED ASSESSMENT AND PLANNING*,†

Richard J. Olson, F. Glenn Goff and Jerry S. Olson

Environmental Sciences Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37830

SUMMARY

Research in energy related assessment and planning at Oak Ridge National Laboratory involves investigating environmental themes at several levels requiring data at appropriate spatial and temporal scales. In the Environmental Sciences Division, we are developing a spatial database for the Eastern United States at the county-subcounty unit level of resolution. The database contains information on terrain, water resources, climate, land use, forest resources, agriculture, wildlife resources, critical natural areas, human population and energy uses. We have defined a spatial hierarchy of metric, geodetic and geopolitical scales as a framework to organizing the data. Spatial units within a hierarchical level serve as building blocks that can be assembled or aggregated to satisfy analysis needs. Building blocks also allow accessing more detailed spatial data by using pointers to information not stored in the database. Uses of the database are related to the capability to cross-reference and integrate information in various subject sectors, utilizing spatial units and temporal periods commensurate with regional themes. An investigation of potential changes in vegetation patterns related to predicted temperature changes from increased atmospheric CO_2 is presented to illustrate an ongoing application of our data resources. Other themes include coal extraction in Appalachia, landscape patterns, habitat and population dynamics of selected biological species, and energy facility siting.

INTRODUCTION

The production of ample clean energy in the United States and other nations of the world requires governments to make choices concerning the development of alternative resources and technologies. Also related to the energy crisis are scientific and public concerns about large scale cumulative effects of agriculture, manufacturing, transportation and urbanization activities on the environment. The increased concern about environmental quality is resulting in an increase of data needs and a proliferation of data gathering. Integrated assessment and planning requires data on various aspects of land, water, air and biological resources available in compatible formats for large geographic regions. Our research is sponsored primarily by the Energy Research and Development Administration (ERDA) and is designed to obtain comprehensive quantitative information on the spatial and temporal distribution of environmental resources and to place these data into a standardized spatial database permitting quick response to identified needs. We are part of the Environmental Sciences Division at Oak Ridge National Laboratory (ORNL) which has an active regional studies program aimed at understanding ecological patterns and processes for integrated assessment of environmental impacts. ORNL researchers are the prime users of our data resources through our collaboration with various projects; however, we exchange and share information with other groups having similar needs and concerns. Our paper describes how we organize and manage large amounts of diverse environmental data to aid in energy related assessment and planning.

The scope and applications for which our database is being used can be illustrated by posing the following questions:

- Where are the optimal areas to locate energy facilities which will have minimal cumulative impact on the region?
- Are there natural communities or agricultural areas located such that SO_2, NO_x or CO emissions may significantly reduce growth or eliminate sensitive species?
- If accelerated burning of fossil fuels increases atmospheric CO_2 concentrations enough to alter temperature patterns, what changes may be anticipated in the distribution of plants and animals?
- Where might biological productivity be great enough to allow renewable organic fuels, such as crop residues or forestry slash, to be used as major sources of energy?
- Where are the areas that possess unique habitat characteristics, e.g., those required by rare and endangered plants and animals? What is the capacity of these areas to support natural populations?
- By comparing areas at different stages of urban or industrial development, can we define the consequences of urbanization and industry on natural plant and animal communities?

As this partial list indicates, most of our research problems extend over large (multi-state) areas and are concerned with assessing longterm changes in ecosystems related to aggregated effects of man's activities.

---

*Research sponsored by the Union Carbide Corporation under contract with the Energy Research and Development Administration.

†Publication No. 901, Environmental Sciences Division, ORNL.

SPATIAL INFORMATION SYSTEMS

Large, automated, spatial information systems are still evolving. They are often called SPATIAL, GEOGRAPHIC, ENVIRONMENTAL or LAND RESOURCE Information Systems. They all entail the manipulation of data associated with locations within large geographic areas. Although all systems attempt to represent real world patterns and variability, each system has its unique methods for accomplishing this within the confines of a digital computer. The data contents of these systems are relatively stable even though the uses of the data may be quite variable. Spatial information systems support decision-making such as assessing impacts, suggesting locations for specific land use or suggesting land uses for specific locations. These characteristics are presented as part of "Information/Data Handling: A Guidebook for Development of State Programs" published this year by the U. S. Department of the Interior.[1] That report, with supporting documents and appendices, represents a fairly complete guidebook for developing spatial information systems even though its objective was to aid states in developing critical natural area programs.

The Information/Data Handling report emphasizes the need for selecting an appropriate spatial scale to collect, store and manipulate data relative to intended applications. Maps have been the traditional media for representing spatial information and serve to illustrate the concepts of scale and resolution. Small scale global maps are valuable in representing major climatic patterns, biomes, etc. whereas large scale 7-1/2 minute quadrangle maps (1:24,000 - USGS) are very valuable in representing local relief, development and potential land use sites. We have developed several ways of expressing this spatial hierarchy of regions (Table 1). The schemes are based on units that are metric (assumes small areas of the earth can be treated as planes), geodetic (size of cells varies depending on latitude) or geopolitical. The regions are matched with map scales commonly used to represent each level. Equivalent relationships were proposed by Goff et al.[2] In each series, the area of the region increases exponentially by a factor of 100.

Regular cell sizes in the geodetic and metric schemes are convenient for computer processing, display and modeling; however, geopolitical units are often more appropriate for data gathering and socioeconomic considerations as well as for action program implementation. Although exceptions exist, as a general relationship, the appropriate spatial cell size for studying regions at one level in the hierarchy is cells at the next lower level. These relationships of scale provide a framework for organizing and relating spatial data, especially as regional studies are broadened to national and international scales.

GEOECOLOGY DATABASE

Our "Geoecology Database" is somewhat unique in attempting to store diverse types of data for large geographic areas. Most databases with broad geographic coverage are narrower in their subject content while those including many subject areas are limited to a single state or smaller area. We have selected the county-subcounty as our spatial unit although we have data at other scales and formats. Extant data have been obtained from national surveys, national networks of monitoring stations, digitized maps and other sources. We currently have over 500 variables stored in files with consistent identifiers and compatible formats for counties in the eastern half of the United States (Table 2).

We use large IBM computers in the batch mode with magnetic tape and some disk storage. The Statistical Analysis System[3] developed at North Carolina State University is currently providing file management and statistical analysis capabilities. The Oak Ridge Regional Modeling Information System (ORRMIS)[4] is being modified to handle county data in addition to its ability to handle hierarchical cell structures, points, lines and polygons. We intend to use ORRMIS when the modifications are completed. The Geographic Data Systems Group at ORNL has developed numerous computer programs to analyze and display spatial data as part of ORRMIS. As an example, to analyze the effects of strip-mining in the Appalachian region, that group aided in relating aerial photography, LANDSAT(ERTS), TOPOCOM(topographic data), known coal reserves and population data to land use, aesthetic, hydrologic and economic impacts for a 7-1/2 minute quadrangle area.[5] Computer-generated cartographic displays, including 3-D perspective drawings, were utilized extensively in presenting the results. The Regional and Urban Studies Information Center (RUSTIC) located in the Energy Division at ORNL maintains a large collection of computerized socioeconomic and energy related files which are available to us and to outside groups.[6] Through these relationships with other ORNL Divisions in data processing, our project is an integral part of the total information and analysis activities related to regional environmental and energy studies at ORNL. The next step of coordination with other ERDA national laboratories is underway.

We are currently standardizing and documenting the data files and assembling software for our database. Standardized files on hand to date include terrain, agriculture, land use, plant and animal species ranges, bird surveys, climate and human population. Files on water quality and quantity, air quality, forest resources, critical natural areas and energy use are being processed for addition to the database. We will be able to establish temporal changes with the database as we add data measured or re-inventoried over time. Currently, most files were gathered during the base period of 1965 to 1974. In addition to adding temporal coverage, we anticipate expanding the geographic coverage to include the western states and to add subject files as required by research projects. The system and data contents were described in more detail in a paper presented at the recent Fifth Biennial International CODATA Conference.[7]

Difficulties encountered in assembling data from diverse sources can be illustrated by describing the development of several subject sectors. National surveys that aggregate data to the county level present the least problems as source data for our database. However, federal agencies collecting data often use different county identifiers and different sets of counties both of which must be resolved as part of reformatting records for the database. Both the 1969 Census of Agriculture conducted by the Bureau of the Census and the 1967 Conservation Needs Inventory (CNI) conducted by the Soil Conservation Service provide land use acreage by county that we are incorporating into our database.

The CNI used field surveys to classify land according to 19 land uses and according to conservation needs (erosion control or drainage) required to improve agricultural capabilities. Unfortunately federal lands, urban areas and small lakes were excluded from the survey. We have supplemented the files with

acreages for these three classes by county but are still attempting to identify specific land use categories of the federal lands which constitute one-third of the nation. Maps and diverse published tables are being used to locate and describe federal land according to the administrative agency, e.g., National Park Service, Forest Service, Bureau of Land Management, Armed Services, etc.

The Census of Agriculture is restricted to farm land surveyed by mail questionaires. Approximately 300 records per county are available on summary tapes from the Economic Research Service. We abstracted 46 crop acreage and yield totals per county to represent major crops. Both this and the CNI survey have some similar categories for crops such as "close grown crops," "row crops," "hay," etc., but acreages from the surveys do not agree. Differences may be due to different classification schemes, two year time difference, sampling differences, or withheld data in the Census of Agriculture related to data privacy considerations. Although both files represent very useful county data, the pecularities associated with each must be documented and considered in using the files.

Developing a climatic subject sector has presented many more difficulties in attempting to characterize the climate for each county. Our desire is to store average seasonal patterns for temperature, precipitation, moisture, winds, radiation and other climatic parameters. The sector is being created from interpolated averages based on Weather Bureau station records, interpolated values from maps in the Climatic Atlas of the U. S. and derived variables such as the Thornthwaite evapotranspiration index. The Weather Bureau's primary mission is short-term weather prediction for populated areas. Consequently the 200-250 first order (most complete climatic records) stations in the east are located at large cities or airports. Temperature and precipitation monthly norms are available for the periods 1931-1960 and 1941-1970 for approximately 1600 stations in the east; however, regions with abrupt climate changes, such as near mountains, have few stations. An unexpected problem encountered in processing the station data was obtaining a digitized file of the latitude and longitude of each station with county identifiers. We ended up manually extracting this information from Weather Bureau publications.

We are developing interpolation algorithms to estimate weather data for all counties. To interpolate values for county centroids, we use mathematical functions that define the characteristics of a hypothetical surface for the digitized points. An equation of a tangent plane for each digitized point is calculated using the closest 3-12 points. Next these equations for points surrounding a centroid are solved to give values for each tangent plane at the centroid; these values are averaged to give the estimated value for the county.

Maps entail several steps in data preparation prior to obtaining county values. We have used an electronic digitizing device to store coordinates of points or isolines on a map in computer readable form. Subsequently these data must be interpolated to counties using the same methods described for station data. Finally the file must be displayed and compared against the original map for editing. The resulting information is limited in accuracy relative to the combined factors of base map resolution and mathematics of interpolation.

The climatic sector has utilized a variety of data sources and analytical tools described above to obtain the information we require to investigate the influence of climate on ecosystems and the role of climate in distributing and diluting pollutants in the environment. We anticipate, as demands for spatially distributed data increase, that mission oriented agencies such as the Weather Bureau will develop capabilities to produce this information from their files in digital form. We also anticipate providing these agencies with feedback as to our research needs which could be met by addition of stations or parameters being measured. We acknowledge that there are deficiencies in our climatic sector but feel that this first iteration may provide guidance to future refinements.

There are 2660 counties in the eastern 37 states ranging in size from 60 to 17,700 km^2, with an average of 1750 km^2. Some advantages of the county as a spatial unit are: well established and quite permanent geopolitical units; common unit for state and federal data collection; uniform in size in the east; often used as management units for environmental policy; boundaries often approximate geophysical features; and county outlines are available in computerized form. Disadvantages include: occasionally names or boundaries change; a few counties are large or heterogeneous; some models use grids of uniform cell size; and independent cities representing some urbanized administrative areas are treated differently relative to data collections. We have addressed the problem of large counties by proposing a set of standard subcounty units (SCU's). By subdividing 79 counties based on size and geographic heterogeneity, the largest subcounty unit in the east was reduced to 6,900 km^2 and the average to 1660 km^2 while the number of SCU's increased to 2800. We are entering some data for these smaller units and using them in cartographic display. Counties in the West are often large and heterogeneous and this limits their usefulness as spatial cells unless a common set of SCU's can be defined.

Data summarized at the county or subcounty scale of resolution can be analyzed to investigate a variety of "macro" patterns and processes. The influence of climate on broad vegetation and crop zones can be effectively described in terms of such analysis. More direct assessment related patterns and processes, such as the influence of temperature regime on the feasibility of development of given energy technology alternatives, can also be addressed using county level data as "building blocks" for the analysis. In this usage, counties can be viewed as sampling "plots" for nation-wide systems analysis.

The Energy Division at ORNL has used county-level studies to identify candidate counties for energy development.[5] These studies located counties based on demand, water availability and other criteria which potentially would support an energy facility. We anticipate refining these predictions based on considerations of local and regional environmental impact. The resulting list of counties would then require additional studies at a finer scale of resolution prior to the time at which a power company or public utility may explore tracts potentially available. Local constraints would narrow the final selection; however, this sequence will result on establishing data files for use in impact statement preparation.

A vast amount of information exists about our environment. The "National Inventory of Biological Monitoring Programs" being conducted at ORNL has canvassed over 3000 biomonitoring and baseline projects in the United States. The "National Index of Data Bases" compiled by Lawrence Livermore Laboratory

contains over 4000 descriptions of databases related to energy and environment. The Ecological Sciences Information Center, the Environmental Response Center and other information centers at ORNL provide bibliographic and state-of-the-art literature reviews which contain valuable numeric data to complement our spatial database. Our own activities, sponsored by both ERDA and the National Science Foundation (NSF) in conjunction with the International Biological Program (IBP), have generated nearly 300 data sets.[8] These diverse data sources are individually and collectively extremely valuable; however, these data are often not readily available, have incomplete documentation or were not designed to be compatible with other related studies. Accumulating these files into a single standardized database would be costly and, because of the uniqueness of each data set, scientifically undesirable. Figure 1 represents the relationships of our database and data sources. We are selecting preferred data sets from these diverse sources based on geographic coverage, subject content, sampling time period, sampling techniques and availability for entry into our database. Even though we do not intend to enter the majority of existing environmental data into our files, we consider these sources of information to be an integral part of the resources available for our research efforts.

The county, or other spatial cell, can be used as an effective repository unit for locating and assembling information about more detailed aspects of the cell. The county record can contain information about the information pertaining to the county -- whether a particular type of data is available within that county, how much data are available, the scale of resolution, dates of data collection, as well as other descriptors of the data. As more data are collected at refined scales, "pointers" to these data can be added to the county records. In this way the county can serve as a reservoir of pointers to more refined data and the county record can be utilized as a repository of information about data that may be stored in auxiliary files, retained in parent agency files, or potentially available in other ways. As an example, we intend to store the location and identifiers of Critical Natural Areas as administered or inventoried by states, The Nature Conservancy, The National Park Service, Experimental Ecological Reserve and others in our database. We will not store detailed information or data from these areas but will obtain the data from the administrative agency if we identify counties in which a user requires the more detailed information. Therefore, our approach to storing selected county-level data is an efficient means of getting to the more detailed, voluminous data that are not stored in the database.

APPLICATIONS

We are attempting to use the hierarchical approach to the development of our spatial database to address issues of regional, national and global significance. An example of anticipated impacts form fossil fuel burning illustrates the use of the database and our approach. All nations have been contributing to a nearly exponential increase of CO_2 in the atmosphere in recent decades. The nature of atmospheric mixing and seasonality causes the worldwide CO_2 increase to show quicker increase and greater seasonal amplitude in the high northern latitudes than in either the tropical or southern hemispheric regions. This is due to atmospheric CO_2 being drawn down by photosynthesis during summer months on the broad northern continents and then being regenerated by respiration during the rest of the year. About half of the amount of CO_2 added from fossil fuel burning has been balanced by increased absorption by the oceans and/or the photosynthetic plant cover of the lands. Forests appear to be the most effective of all ecological systems in exchanging and storing carbon as biomass. However, this capacity for increased absorption is being reduced by the clearing of forests. The burning and indirect oxidation (decomposition by microorganisms) of carbon pools formerly associated with forests are contributing to an increase in atmospheric CO_2 level.

Modeling of the global carbon balance by several groups leaves little doubt of drastic increases, possibly doubling, of CO_2 in the next century unless energy technology turns away from oil, gas and coal with the same abruptness that we are turning to these fossil fuels. There is some uncertainty about the degree of warming that could be expected per doubling of CO_2, with the possibility that particulate pollution might either partly counteract or conceivably exacerbate the changes. But recent assessments (e.g., Baes et. al.)[9] leave little doubt that climate change will be significant and spatially variable over the coming decades. Ground investigations need to be initiated to document and predict shifts in life and crop zones resulting from climatic changes. Systematic monitoring of trends in forest or other ecological cover need to be expanded to include tropical regions and other areas where data are currently limited in quantity and quality. The analysis of satellite data, with associated aircraft and ground truth calibration, is of high priority for supplementing more traditional data sources to aid in improving our knowledge of the carbon cycle.

At present, our database is being used to investigate productivity, decomposition and mineralization rates in the eastern deciduous forest to aid in understanding the carbon cycle. Carbon pools, photosynthesis rates and respiration rates are being determined for the vegetation. Landscape diversity as related to climate, terrain and man's conversion of natural areas to agriculture and urban uses, is also being studied. Knowledge of relationships of vegetation and climate is necessary to predict temporal changes in the spatial patterns of vegetation and crops under various projected climatic shifts. Having terrain, forestry, agriculture and climate variables available on common spatial units and in compatible formats allows ready cross-reference and combining of data from these various subject sectors.

Using the established requirements for a crop (e.g., temperature, moisture, soil, and associated variables for growing corn), we are attempting to predict changes in the growing zone and differential productivity within the growing zone resulting from temperature changes induced by factors such as increased atmospheric CO_2. County corn yields can be adjusted to reflect the predicted seasonal temperature changes with some counties potentially being unable to produce corn and others switching from other crops to corn production. These switches in the spatial pattern of the "corn belt" along with other growth modifications would be used to characterize a new hypothetical corn region. The region's character, using counties as spatial cells, will be based on the aggregated properties of the cells. Additional regional characteristics for the larger area may emerge that are not readily apparent from any one cell in the region. This corn region, with its emergent properties and larger geographic coverage, will then become a spatial unit to be used in studies at the national or global level. This example shows how the hierarchical approach is being used to aggregate spatial units in an environmentally meaningful way to integrate studies conducted at various regional scales.

DISCUSSION

In October 1975, a week-long workshop on energy related regional analysis and assessment was held in Oak Ridge, Tennessee, with representation from most of the ERDA National Laboratories. In consideration of ERDA's interest and role in development of such a database, the following statement was drafted by the task force on Regional Environmental Data Bases:

Recognizing:

(1) That decisions regarding energy research and development should be made on a rational basis with due regard for the optimum utilization of all human and natural resources, as well as for minimum desecration of land and minimum degradation of environment;

(2) That such rational decisions require a base of spatial and temporal data relating to human needs and conditions and to the environmental resources fulfilling these needs;

(3) That interagency cooperation and coordination in such an information base is necessary to avoid duplication of inventory efforts, to insure adequate "quick response" availability of data on the multiple subject sectors that must be considered in responsible rational resources development, and to insure optimal design of agency inventory, planning and program implementation efforts;

(4) That ERDA-DBER, through the National Laboratories, is uniquely qualified to perform the collation and synthesis of selected data from the diverse array of data presently available; and to determine data needs and mediate in the design and implementation of future inventory efforts to insure maximal relevancy of data to energy development, planning and assessment needs.

The Task Force on Regional Environmental Data Bases recommends that ERDA-DBER foster the design and development of a national spatial data base for energy resources planning, assessment, and analysis. Development of this data base should be undertaken as a cooperative interlaboratory effort with appropriate division of responsibility and resources among the laboratories.

With respect to coordination with other agencies, we recommend a "distributed data base" concept and approach with extensive data on natural and human resources as well as environmental conditions residing in the responsible agency. For ERDA analysis and assessment, the database should be developed as a spatial data management system with quick response capability, high exportability (simple structure), and an optimal level of spatial and temporal standardization, for "preferred data" to be assembled and synthesized from the "distributed data base" of other agencies and sources.

Although this set of recommendations does not represent official ERDA policy, it does indicate the central role of energy related assessment and planning. Whereas in many other agencies one clear theme arising from a traditional discipline-oriented mission renders consideration of all other types of information clearly subordinate, in energy related assessment and planning, ERDA (possibly along with the Environmental Protection Agency and the Council on Environmental Quality) must balance concerns for all aspects of regional development.

We have attempted to define a spatial hierarchy as a key in organizing and using the vast amount of environmental information available for energy related assessment and planning. These hierarchical relationships can guide in aggregating data to larger regions or in "zooming" into smaller areas. The spatial cells within a level of the hierarchy represent building blocks which can be assembled into regions according to various needs. Each block can contain selected data for key environmental parameters to allow analysis at that level of resolution. The blocks can also contain pointers to other sources of information not stored in the database. This scheme has allowed us to build an environmental database utilizing counties as geopolitical spatial cells even though this level of resolution does not contain all the detailed data for studies at local levels. Using county cells, we are investigating ecological patterns and processes of regions covering several states relative to the impacts of energy development. Our confidence in this scheme lies in being able to identify potential areas for more detailed study and to access additional information using the hierarchical relationships and pointers.

Construction of spatial databases, such as ours, can bring about substantial integration and coordination among agencies and groups having specific or general interest in the themes addressed by a database. Most of this integration occurs in using the database but a large part may also result from designing the database, defining its subject content and assembling the data files. The database can become an important project management tool by bringing about more effective communications between the individuals with diverse backgrounds and interests involved in data acquisition and synthesis. This is especially true as international groups such as those sponsored by NATO cooperate in building databases to study global processes.

We have not discussed many of the technical decisions involved in developing a spatial database such as georeferencing, digitizing, file structures, cartographics and many others. Neither have we discussed the many "people" problems associated with automated systems which may outweigh the technical problems in determining the success of such systems. In light of the worldwide needs for regional environmental data, it is imperative that groups such as AGARD continue to participate in planning, leading, organizing and controlling the development of spatial information systems.

REFERENCES

1. Department of the Interior, *Information/data handling: A guidebook for development of state programs (draft)*. Office of Land Use and Water Planning and U. S. Geological Survey Resource and Land Investigations Program. 1975. 303 pp.

2. F. G. Goff, F. P. Baxter, and H. H. Shugart, Jr., *Spatial hierarchy for ecological modeling*. Oak Ridge National Laboratory. EDFB Memo Report 71-41, 1971. 12 pp.

3. A. J. Barr, J. H. Goodnight, J. P. Sall, and J. T. Helwig, *SAS-76, A user's guide to the statistical analysis system*. North Carolina State University, Raleigh, North Carolina. 1976.

4. R. C. Durfee, *ORMIS, Oak Ridge Regional Modeling Information System, Part 1*. Oak Ridge National Laboratory. ORNL/NSF/EP-73, 1973. 103 pp.

5. *Energy Division Annual Progress Report, Period Ending December 31, 1975*. Oak Ridge National Laboratory. ORNL-5124, 1976. 209 pp.

6. A. S. Loebel, D. F. Burch, and N. S. Malthouse, *Regional and urban studies socioeconomic data bases at Oak Ridge National Laboratory*. Oak Ridge National Laboratory. ORNL/TM-5182, 1975. 57 pp.

7. R. J. Olson and F. G. Goff, *Development and applications of a regional environmental data base for southeastern United States*. Presented at Fifth Biennial International CODATA Conference, Boulder, Colorado, June 28 - July 1, 1976.

8. J. A. Watts and M. L. Johnson (eds.), *Abstracts US-International Biological Program ecosystem analysis studies*. Vol. IV, No. 4. Oak Ridge National Laboratory. US/IBP-76/1, 1976. 300 pp.

9. C. F. Baes, Jr., H. E. Goeller, J. S. Olson, and R. M. Rotty, *The global carbon dioxide problem*. Oak Ridge National Laboratory. ORNL-5194, 1976. 70 pp.

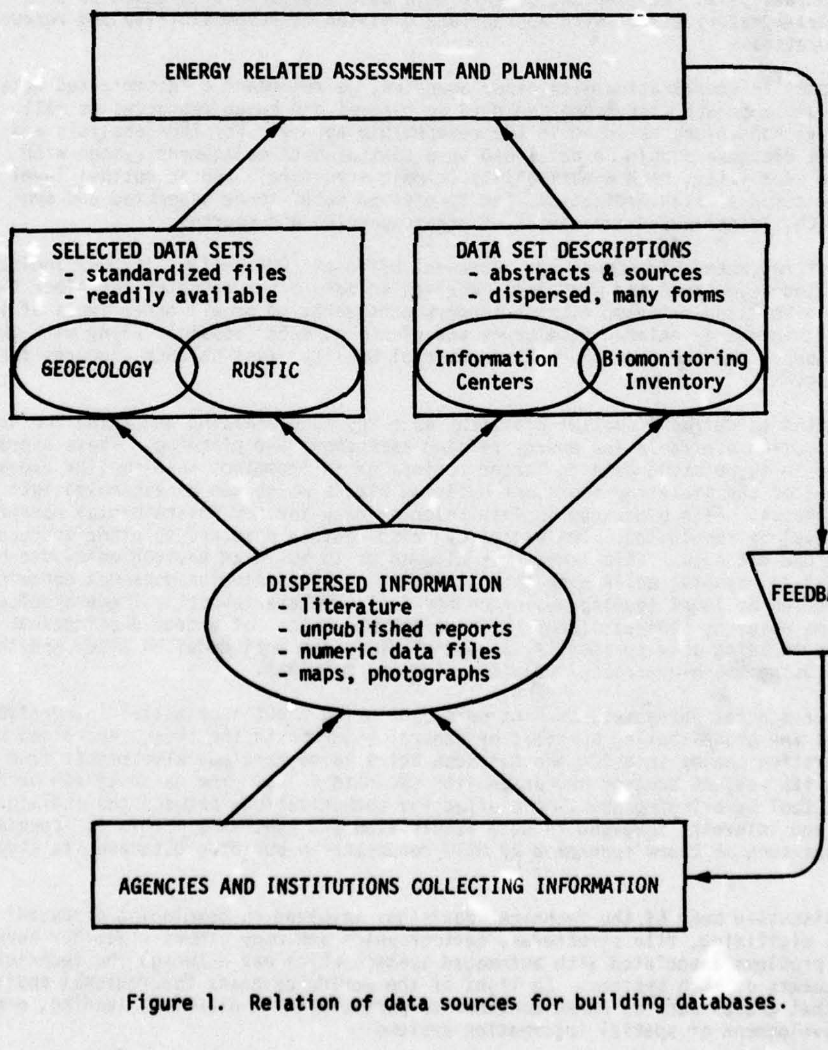


Figure 1. Relation of data sources for building databases.

Table 1. Spatial hierarchies of regions

Level	Area km^2	Metric side-km	Geodetic side	Map scales	Geopolitical typical unit	Geopolitical area-km^2
R1 - World				Smaller scale	World	5.1 x 10^8
	10^8	10000	90°	1:20,000,000		
R2 - Mega region					Continents (7) Nations (127)	2.1 x 10^7 1.2 x 10^6
	10^6	1000	10°	1:2,000,000		
R3 - Macro region					States & provinces of N. Amer. (126)	1.8 x 10^5
	10^4	100	1°	1:200,000		
R4 - Meso region					Counties in East U.S. (2660)	1.8 x 10^3
	10^2	10	7 1/2'	1:20,000		
R5 - Micro region					Townships	0.9 x 10^2
	10^0	1	30"	1:2,000		
R6 - Macro site					Watersheds, Research plots	
	10^{-2}	0.1	3.75"	Larger scale		

Table 2. Geoecology database contents as of July 1976

Data set description	Year and source	Area	Variables /record	Records /county
County totals				
Agricultural crop acreage and production	1969 Ag. Census	USA	46	1
Livestock and poultry inventory and sales	1969 Ag. Census	USA	28	1
Land use	1967 Cons. Needs Inv.	USA	33	1
Land capability	1967 Cons. Needs Inv.	USA	35	1
Land use by land capability class	1967 Cons. Needs Inv.	USA	28	12
Population by sex and five year age classes	1970 Census	USA	45	1
County codes, names, area and centroids	1970 FIPS	USA	14	1
Station data				
Weather station locations	1970 Weather Bureau	East	10	2.6
Normal monthly maximum temperature	1931-60 Weather Bureau	East	19	.1
Normal monthly minimum temperature	1931-60 Weather Bureau	East	19	.1
Normal monthly precipitation	1931-60 Weather Bureau	East	19	.9
Normal monthly heating degree days	1931-60 Weather Bureau	East	19	.6
Normal monthly average temperature	1931-60 Weather Bureau	East	19	.6
Evapotranspiration index	Thornthwaite Equation	East	19	1
Moisture index	Thornthwaite Equation	East	19	1
Breeding bird survey routes	1966-74 Fish & Wildlife	East	25	.5
Breeding bird survey counts	1966-74 Fish & Wildlife	USA	25	30.6
Digitized map data				
Potential natural vegetation	1969 Küchler's Map	East	8	2.2
Tree species ranges	1971 Little's Atlas	USA	6	39.5
Mammal species ranges	Various sources	S.E.	6	
Land surface form	1963 Hammond's Map	East	23	1
Underlying bed rock	1970 National Atlas	East	13	1
Principal soils	1967 SCS	East	9	2.1
Glacial deposits	1970 National Atlas	East	19	.4
Karst terrain occurrences	1970 National Atlas	East	11	.3

Evaluated Numerical Data for the SST and Chlorofluorocarbon Problems:

A case study of how to help the engineer and the modellers

David Garvin and Robert F. Hampson

Physical Chemistry Division
Institute for Materials Research
National Bureau of Standards
Washington, DC 20234

SUMMARY

Mathematical simulations (models) of the stratosphere have been a primary tool in the analysis of potential pollution of the stratosphere from high flying aircraft, rockets and the release of organic chlorine compounds at the surface of the earth. These models require large amounts of numerical data about the meteorology and chemistry. The chemical data comes from laboratory measurements and must be interpreted and made available to the user community in an understandable form and, preferably, as recommended values.

How chemical data were supplied to the Climatic Impact Assessment Program of the U.S. Department of Transportation is described here with emphasis on the role played by the Chemical Kinetics Information Center of the National Bureau of Standards. This included planning, identification of needed measurements and available measurements, determination of the needs of users, evaluation of data, interpretation of results for non-specialists and distribution of tables of rate data. This type of role is suitable for an information analysis center in any large scale interdisciplinary program.

1. Introduction

This paper recounts the roles a scientific discipline oriented information analysis center has played in a multi-technology program which had a clearly defined applied mission. These roles are: advisor to the program managers, coordinator of experimental studies, evaluator of data and communicator with the user groups. Similar roles are appropriate for other data centers, they are desirable for efficient program management and can be of great benefit to the centers.

The center is the Chemical Kinetics Information Center (CKIC) at the U.S. National Bureau of Standards (NBS). The Program with which it has interacted is the Climatic Impact Assessment Program (CIAP) of the U.S. Department of Transportation. These are described below, briefly, to provide a background for discussion of the interaction.

1.1 The Center. The Chemical Kinetics Information Center is part of the National Standard Reference Data System which is managed by the Office of Standard Reference Data at NBS. The Center was established in 1963 and is located in the NBS Physical Chemistry Division. The location is important because the Physical Chemistry Division has extensive experimental programs in gas phase reaction kinetics, photochemistry and ultraviolet spectroscopy. Thus, expertise is readily accessible for most problems faced by the Center.

The Chemical Kinetics Information Center collects, indexes and abstracts published papers and reports on measurements of rates of chemical reactions. It publishes tables of data and bibliographies and also responds to requests for information. To date it has accumulated a data base of some 26,000 papers, mostly in machine readable form. This data base is the raw material used for its other operations. To be useful for problems of current interest it is mandatory that the data base be kept up to date. This combination of technical expertise plus publication and distribution capabilities is ideal for supporting an applied problem.

1.2 The Program. The Climatic Impact Assessment Program (CIAP) was established in 1970-1971 by direction of the U.S. Congress and was charged with determining the potential ecological impact of propulsion effluents from fleets of supersonic transports flying in the stratosphere. The task of CIAP was to reach scientific conclusions that could be used for establishing technical and operational standards for future air travel that would assure a chosen level of atmospheric quality. For its work the program drew on ten U.S. federal departments and agencies and seven governmental organizations in other countries. Some one thousand scientists and engineers contributed to the technical outputs.

It was necessary to collect data on and solve stratospheric problems involving chemistry, trace constituent concentrations, meteorology, radiation flux, engine emissions, biological effects of ultraviolet light (on plants, organisms and man), climatic changes (and their effects on crop production), and economic projections of likely SST fleet sizes [1]. The results of these studies were published in a "Report of Findings" [2], six technical monographs [3] and the proceedings of four

interdisciplinary conferences [4], as well as many papers published separately by the contributors.

In addition to laboratory experiments, atmospheric measurements, field tests and modelling, a program as diverse as CIAP requires a substantial organization to handle coordination of the work, publications and decisions about what to do. Interpretation of technical results is needed by persons at all levels of this hierarchy.

As CIAP came to an end, another problem in stratospheric chemistry arose. This is the effect of chlorocarbon emissions on the ozone layer. The programmatic organization within the U.S. to study this problem is more fragmented than for the case of the SST. The models are slightly different because the source is at ground level, but the technical requirements for chemical data remain the same. The only change is that an additional set of chemical reactions must be incorporated in the models. The descriptions of data needs and information analysis center actions that follow emphasize CIAP but are applicable to the new problem as well.

2. Stratospheric Chemistry and Data Needs

The stratosphere is a region of intense chemical activity. For the present discussion the details are not important. It is sufficient to note that the chemistry involves 50 or so molecules and free radicals, present at the part per million concentrations level or less, that can interact in 150 to 200 elementary chemical reactions and photodissociation processs. The principal interest has been in how these reactions control the amount of ozone in the stratosphere, thereby determining the intensity of ultraviolet light that reaches the earth's surface.

In order to study this problem mathematical simulations (models) are used. They combine the chemistry and meteorology in a prediction of concentration levels of trace substances as a function of geographic position and altitude.

The models require large amounts of numerical input data. The results of the modelling can be sensitive to the chemistry included and the numerical parameters concerning them. Because important decisions may be based on the outputs of the models it is necessary to have high quality input data.

For simple chemical models (minimal meteorology) the following types of data are needed:

(1) The chemical mechanism, i.e., the specific chemical and photochemical processes that should be included to make the outputs meaningful.

(2) The rate constant (as a function of temperature) for each elementary chemical reaction in the mechanism.

(3) The optical absorption cross section for each molecule that absorbs visible or ultraviolet light.

(4) The efficiency of photodissociation (quantum yield) for each molecule that can be decomposed by visible or ultraviolet light.

(5) The intensity of solar radiation as a function of altitude, geographic position, time of day and season.

(6) The temperature and pressure regimes under which the chemical processes occur.

Our concern in the remainder of this paper will be with data derived from laboratory measurements, items (1) through (4), above, who supplies them, who uses them and how a data center can organize this transfer of information.

3. Measurers and Users of Data

The participants in a program can be classed either as measurers (suppliers of data) or users. Some persons are members of both groups, but there is rarely any doubt about their function at a particular time.

3.1 Measurers. The measurement community for atmospheric chemistry consists of many small research groups (two to six persons each) in university, government and industrial laboratories. Their members are predominately chemical kineticists and photochemists. They may be identified easily: they publish scientific papers devoted to explaining fairly simple physical and chemical systems. They produce data of the first four types listed in the previous section and also develop methods of measurement. For CIAP there were chemical research groups in over 100 institutions.

The information analysis center is part of this community. It is relatively easy to establish rapport with the pertinent research groups, and it is important to cultivate them: they provide the data retailed by the center to the users.

3.2 Users. The user community is more diverse but may be smaller and more centralized than the measurers. One major task of a center is to find and to keep in contact with this community.

The users of chemical data for modelling of the stratosphere fall into several classes as described below. Probably there will be analogous classes for any applied program.

(a) Modellers. In the CIAP program there were at least 10 independent groups making mathematical simulations of the stratosphere. Such groups include meteorologists, mathematicians, computer specialists and often, but not always, chemists. Their concern is with the dynamics of the atmosphere. They should not have imposed upon them the burden of selecting chemical kinetics data. These groups are the classical primary users in a mission oriented program.

(b) Combustion engineers. This group is composed of persons concerned with engine tests and with the short time scale dispersal of effluents. Often they model their experiments and usually they use chemical inputs, although these may be less extensive than for other classes of users. The size of this class is unknown, but it was not large.

(c) Chemists. A surprising, major class of users of kinetics data is composed of the chemists making laboratory measurements, i.e., the measurement community. These persons must be considered in the planning of an applied program, they are heavy users of data.

Data are used by chemists in several ways. One is in the design of experiments: the measuring techniques must match the expected time scale and the concentration levels. Another is in interpretation of results and reduction of data. Some overall chemical change is measured in an experiment. Very often five or more elementary reactions combine to produce this change. The rates of the elementary steps are wanted as input for the models. But usually one can extract information on only one or two of these from an experiment. The rate constants remaining reactions are taken as known. Thus, tabulated data can provide standardized base points. A third application is qualitative. The data are used in developing explanations (mechanisms) of atmospheric behavior. Tabular data can indicate which processes could or could not be important. This is a screening function.

(d) Regulatory agencies. There is a small class of persons in regulatory agencies, ad hoc study groups and international committees. They need to understand the chemistry and predictions based on it, want information that can be used in developing methods of monitoring or use it in setting regulations or standards.

(e) Other applications. Some technologists working on problems far removed from stratospheric chemistry found the chemical data to be useful. Such spin-off may never benefit the program for which the data were developed, but it is a source of ideas for future information analysis center activities.

4. Interactions between a program and an information analysis center

Many steps must be taken before it is possible to provide modellers with reliable data. Much more than information transfer is necessary. The needs of the users must be determined and matched with existing data or possible experiments. Laboratory measurements must be made. They must be assessed for validity and selections made among discordant data. Then the material must be put into a form acceptable to the users. Finally it must get to them and be used.

The role of the Chemical Kinetics Information Center in this process is described below as an example. That role developed by trial and error. Interaction with the users and with the management of CIAP shaped it. Both groups pointed out important things that should and could be done.

Four interactions are identified and described below. They occurred throughout the course to the program. A chronological account is not given except where it helps explain a particular interaction.

4.1 The center's role as program advisor. Early in the program the CIAP managers determined that a detailed study of stratospheric chemistry would be necessary. They then asked the following basic questions:

(a) What chemistry is important and what can be ignored?

(b) What is known about these chemical reactions; for which are there usable data and for which are measurements needed?

(c) Which laboratories have the capability for making the measurements?

These questions were answered by kineticists and photochemists in the NBS Physical Chemistry Division, using the data base accumulated by the Kinetics Center. The questions were difficult, had to be answered quickly and specifically. Many guesses had to be made.

Later, as new technical ideas were put forward and new measurements made, program managers turned to the Center for interpretation and assessment of impact. Editors sought answers to questions about nomenclature, units and writing chemical reactions.

This advisory role did much to establish the usefulness of the Kinetics Center in the eyes of the program managers. It is a service that information analysis centers can and should offer to applied programs. Not only does it pay off but it keeps a center on top of current problems.

4.2 The role as coordinator. The technical output of CIAP was organized around a series of six monographs [3], the first of which is on the natural stratosphere and contains a chapter on the chemistry. This was written by a panel of chemists, mostly kineticists, including two members of the Center. The first draft was written at a week-long meeting in November 1972. During this meeting the panel also prepared a list of chemical reactions to be considered in modelling the stratosphere and a table of rate constants and photochemical data. Another meeting was held a year later. Two major revisions of the chapter were made prior to publication.

The Chemical Kinetics Information Center provided technical and editorial support for this Chemistry Panel. For the initial meeting it supplied a library of papers on kinetics. For each of the drafts it coordinated, edited and occasionally retyped the contributions of the panel members. It prepared the bibliography. The Center took charge of the tables of rate data set up by the Panel, revising and expanding them throughout the course of the program (see section 4.3).

This type of coordination is ideal for an information analysis center. It is an interaction with its home community - the data generators. The Center provided this measurement community with information on the state of the art, the best data to use and lists of measurements that were needed. In return it received a steady flow of reports on new work. For the CIAP Chemistry Panel the Center did all those jobs that annoy authors and at the same time attempted to assure a measure of uniformity and coherence in the final publication.

4.3 The role as data evaluator. Before data are used in an applied problem the wrinkles in the laboratory results need to be smoothed out. The user wants a number, preferably with some statement of its reliability, not a group of numbers from which to select. Very often a user is not an expert in the discipline that supplied the data and should not attempt to interpret the primary research reports.

Codification and evaluation of data are the prime functions of an information analysis center and are the jobs it understands best. They should be central to the role the Center plays in any applied program.

Data evaluation is much more than simply producing a set of reliable numbers. It includes an analysis of the demands of users in order to determine their actual needs and how these mesh with the available data. It also requires careful attention to how the associated applied program evolves in order to keep the evaluation work in line with changing emphases. And above all, the evaluators must know about the latest experimental work. New data must be obtained, analyzed and distributed quickly. Otherwise the more knowledgeable users will lose faith in the recommended numbers.

The needs of CIAP for chemical data were extensive and varied. They were met in several ways. First, a survey was made of papers on stratospheric chemistry to determine which reactions were important and what new measurements were needed either to improve the accuracy or to fill gaps in the data base. Second, a workshop for chemical kineticists and atmospheric modellers was held at NBS in the Fall of 1972. This provided a forum for modellers to describe their needs and experimentalists to outline measurements that could be made in the near future. Third, as indicated in the previous section, an initial table of rate data was developed by the CIAP Chemistry Panel. This got the modellers off the ground. Fourth, a data evaluation program was mounted at NBS to extend this initial effort and keep it up to date. It involved scientists both at NBS and in other laboratories, the results being compiled and distributed by the Center. Fifth, this evaluation activity was supplemented by compilations of new data, not yet evaluated.

All of this material was distributed by the Center in a series of reports [5]. The final tables were published in CIAP Monograph No. 1 [3] and made available to the general public as NBS Technical Note 866. Talks were also given at CIAP Conferences [4] in order to keep users aware of new developments. The communications component of evaluation work is surprisingly large.

4.4 The role as communicator. It is not enough to evaluate data. The recommended values must reach the users and reach them in time to be of use. It may be necessary to persuade them to use the data. It also may be desirable to check on how the data actually are used. This means an active communications plan.

The plan for CIAP was very simple. It was to get the tables of data to anybody who had a use for them. There was no preset distribution list, there were no limitations. Any organization that could be identified as having an interest in atmospheric chemistry was put on the distribution list. Published papers and attendance lists of meetings were useful sources. The CIAP newsletter advertised the tables. Requests for tables were honored without question. Over the course of the program the distribution list grew from 250 to 850 persons.

This policy of saturating the market paid off. First, most of the modellers began to use the same input data with the result that their results could be more easily intercompared. Second, the laboratory scientists began to see which data were well established

and which they should measure. Third, a large number of people not formally part of CIAP became aware of that program. Fourth, a feedback of ideas from all sources enriched the studies. And finally, it established the Kinetics Center as the source to query when new data were needed.

In retrospect, the feedback was an important factor in keeping the Center on top of its job. Modellers indicated needs for new data and told what effects certain reaction rates had on their predictions. These facts indicated directions for new work. Laboratory scientists did the same, contributed their measurements in advance of publication and criticized the recommended values. This made the tables current and provided quality control.

Another communications role was an attempt at quality control of the models. This consisted of monitoring the chemical data used by the modellers. Each group was encouraged to send lists of its input data to the Kinetics Center. These lists were checked for currency and reasonableness. Errors were corrected and suggestions were made about the chemistry and about the reliability of the numbers being used. To all appearances this service was appreciated by the groups that took advantage of it. From the viewpoint of the program the service improved consistency among the models without forcing it by fiat.

5. Concluding Remarks

An information analysis center has two things to offer to a mission-oriented program. They are technical expertise in a scientific discipline and experience in information transfer. The first enables it to coordinate and interpret the laboratory measurements used in the solution to the problem. Also, an existing Center with extensive data files can respond quickly in the early stages of a program when important decisions will be made. The second simplifies the service roles of communicator and advisor: techniques used in the past can be brought to bear on a new problem, they need not be reinvented.

The four roles of advisor, coordinator, evaluator and communicator, described here, are not specific either to stratospheric chemistry or to chemical kinetics. They should be applicable to most applied problems that require substantial scientific input. Any information analysis center should recognize ways it can play these roles and use them to make it a focal point in a mission-oriented program.

References

[1] This paragraph is drawn largely from the Executive Summary in reference [2].

[2] Grobecker, A. J.; Coroniti, S. C.; Cannon, R. H. Jr., "Report of Findings. The Effects of Stratospheric Pollution by Aircraft" DOT-TST-75-50 (December 1974) National Technical Information Service (NTIS), Springfield, VA 22151.

[3] CIAP Monograph Series, Grobecker, A. J., Editor in Chief (available from NTIS, Springfield, VA).

No. 1. Reiter, E., et. al., "The Natural Stratosphere of 1974," DOT-TST-75-51 (September 1975).

No. 2. English, J. M., et. al., "Propulsion Effluents in the Stratosphere," DOT-TST-75-52 (September 1975).

No. 3. Robinson, G. D., et. al., "The Stratosphere Perturbed by Propulsion Effluents," DOT-TST-75-53 (September 1975).

No. 4. Leith, C. E., et. al., "The Natural and Radiatively Perturbed Troposphere," DOT-TST-75-54 (September 1975).

No. 5. Nachtwey, D. S., Ed., "Impacts of Climatic Change on the Biosphere. Part 1 - Ultraviolet Radiation Effects," Bartholic, J., Ed., ibid, "Part 2 - Climatic Effects," DOT-TST-75-55 (September 1975).

No. 6. d'Arge, R. C., et. al., "Economic and Social Measures of Diologic and Climatic Change," DOT-TST-75-56 (September 1975).

[4] Barrington, A. E., Ed., "Proceedings of the Survey Conference on the Climatic Impact Assessment Program," February 15-16, 1972, DOT-TSC-OST-72-13. (September 1972) (Available from NTIS, Springfield, VA., as are the others listed below.)

Broderick, A. J., Ed., "Proceedings of the Second Conference on the CIAP," November 14-17, 1972, DOT-TSC-OST-73-14 (April 1973).

Broderick, A. J., and Hard, T. M., Eds., "Proceedings of the Third Conference on the CIAP," February 26-March 1, 1974. DOT-TSC-OST-75-15 (November 1974).

"Proceedings of the Fourth Conference on the CIAP," February 1975, in press.

[5] *Hampson, R. F., Ed., Brown, R. L., Garvin, D., Herron, H. T., Huie, R. E., McKinley, J. D., and Tsang, W., Chemical Kinetics Data Survey I. Rate Data for Twelve Reactions of Interest for Stratospheric Chemistry, NBS Report 10 692 (Jan. 1972).

Hampson, R. F., Ed., Braun, W., Garvin, D., Herron, J. T., Huie, R. E., Kurylo, M. J., Laufer, A. H., Okabe, H., Tsang, W., Chemical Kinetics Data Survey II. Photochemical and Rate Data for Fifteen Gas Phase Reactions of Interest for Stratospheric Chemistry, NBS, Report 10 828 (April 1972).

Garvin, D., and Gevantman, L. H., Chemical Kinetics Data Survey III. Selected Rate Constants for Chemical Reactions of Interest in Atmospheric Chemistry, NBS Report 10 867 (June 1972).

Johnston, H., and Garvin, D., Working Papers for a Survey of Rate Data for Chemical Reactions in the Stratosphere, NBS Report 10 931 (Oct. 1972).

Hampson, R. F., Ed., Braun, W., Brown, R. L., Garvin, D., Herron, J. T., Huie, R. E., Kurylo, M. J., Laufer, A. H., McKinley, J. D., Okabe, H., Scheer, M. D., Tsang, W., and Stedman, D. H., Survey of Photochemical and Rate Data for Twenty-Eight Reactions of Interest in Atmospheric Chemistry, J. Phys. and Chem. Ref. Data, 2, 267-312 (1973).

Garvin, D., Ed., Chemical Kinetics Data Survey IV. Preliminary Tables of Chemical Data for Modelling of the Stratosphere, NBSIR 73-203 (May 1973).

Garvin, D., Ed., Chemical Kinetics Data Survey V. Sixty-Six Contributed Rate and Photochemcial Data Evaluations on Ninety-Four Reactions, NBSIR 73-206 (May 1973).

Hampson, R. F., Ed., Garvin, D., Herron, J. T., Huie, R. E., Kurylo, M. J., Laufer, A. H., Okabe, H., Scheer, M. D., and Tsang, W., Chemical Kinetics Data Survey VI. Photochemical and Rate Data for Twelve Gas Phase Reactions of Interest for Atmospheric Chemistry, NBSIR 73-207 (Aug. 1973).

Garvin, D., and Hampson, R. F., Eds., Chemical Kinetics Data Survey VII. Tables of Rate and Photochemical Data for Modelling of the Stratosphere, NBSIR 74-430 (Jan. 1974).

Watson, R. T., Chemical Kinetics Data Survey VIII. Rate Constants of ClO_x of Atmospheric Interest, NBSIR 74-516 (June 1974).

Hampson, R. F., and Garvin, D., Eds., Chemical Kinetics and Photochemical Data for Modelling Atmospheric Chemistry, NBS Tech. Note 866 (1975).

---

*NBS Reports and publications in the *NBSIR series that are still in print may be obtained* from the Chemical Kinetics Information Center, NBS, Washington, DC 20234.

HOLOGRAPHIC DATA STORAGE AND RETRIEVAL SYSTEM

by
Fred N. Haritatos, IRDE
Jack D. Petruzelli, IRAP
ROME AIR DEVELOPMENT CENTER
Griffiss Air Force Base NY 13441
USA

SUMMARY

The advent of computer technology has resulted in the ever-increasing need to collect, process and store digital information. Until recently conventional storage techniques, through evolutionary product improvements, have been able to keep pace with this growth. However, a point has been reached where further improvement is becoming technically and economically impractical. The United States Air Force has recognized the limitation of conventional approaches and has actively pursued development of alternate storage techniques. As an outgrowth of this commitment, the Rome Air Development Center has developed two mass data storage and retrieval systems which will have a significant impact on present data processing operations. These systems have been designated Human Readable Machine Readable (HRMR) Microfilm Mass Memory System and Wideband Holographic Recorder/Reproducer.

As world events become more intimately related to our nation's security, it becomes more important to maintain large reserves of reference information to accurately and rapidly assess and respond to crisis situations. From all available sources, information is collected and processed by a network of sophisticated data handling systems in support of intelligence, command and control, logistics and other operational requirements. In each application, information storage and retrieval comprises the important resource, with computer systems serving to process and provide accurate, usable information in the right place at the right time. Such applications have stimulated a demand for larger, cheaper, and faster data storage and retrieval capabilities.

To meet this need a number of manufacturers have stepped up efforts to produce new mass storage systems. In fact, recent studies have projected the market potential for such systems to be between \$500 million and \$1 billion. As a result of the increased R&D activity, several advances have taken place; the most dramatic of which have been in electro-optic data storage.

Communications over long distances have been greatly improved by the ability of satellites to rapidly transmit gigabits of data between widely separated points. Satellites have also gained in popularity as a means of monitoring and managing our dwindling natural resources. For example, the Landsat satellite, which now orbits the earth, sends back daily information on waterflow, agriculture, geography, land use, fishing grounds and mineral resources. Repeat observations of the same geographic area are normally scheduled days apart. Also, analysts and scientists require comparative data covering many weeks, months, or years. Consequently, ground data stations must possess the capability for reliably storing and retrieving large volumes of analog and digital data over long periods. Capture and retention of this high volume, highly volatile information has been made possible largely through the development of advanced wideband recording methods.

Today, magnetic tape is commonly employed to satisfy these large data storage and retrieval requirements. Many automatic data processing centers within the Government have libraries containing over 20,000 reels of magnetic tape. Using this storage medium, data can be read processed and written using standard computer equipment. Magnetic tape, however, does have two significant problems that limits its further application. First, it is an expensive storage medium, particularly in the context of large data storage applications. A reel of tape typically costs a minimum of ten (10) dollars, while instrument-quality tape can exceed one-hundred (100) dollars. There is also the expense associated with maintaining the integrity of the tape library. Due to the tape's erasable nature, stored digital data will degrade through use and storage. Magnetic tape is also highly susceptible to spurious electric and magnetic fields which are especially severe in most military environments. This makes tape vulnerable to both data loss and alteration through either accidental or intentional tampering.

The Rome Air Development Center (USAF) has recognized the problems of utilizing magnetic tape for a number of years, and an alternative mass data storage concept, using an optical recording technique, has been developed. This approach has been designated the Human Readable Machine Readable (HRMR) Microfilm Mass Memory System. The system is capable of recording in near real-time both human readable information and its total digital equivalent on a common film format. The significant feature of this concept is that the machine readable data is stored in an unused part of the film chip in holographic form.

One of the reasons for selecting holography is its high information storage capacity. In-depth analyses and experimental breadboard models have demonstrated a capability for reliably storing and retrieving digital data at a conservation density of 250 thousand bits per square centimeter. Compared to conventional external computer memory storage, a significant reduction in storage can be achieved through holographic recording techniques.

In an effort to validate the concept of holographic data storage, a system was designed, fabricated and delivered to RADC in May 1973. The equipment shown in Figure 1 has been actively used for in-house and contractual experimentation. The system consists of the following modules: (1) microfiche recorder; (2) subsystem controller/computer (PDP-11/15); and (3) microfiche reader. Input devices include magnetic tape, document digitizer and modem coupler. Output from the system is in the form of HRMR microfiche, hardcopy from the CRT display, magnetic tape or digital signal to a host computer via a modem unit. System access and control is through the CRT display terminal.

The HRMR system is capable of producing any of three microfiche formats as depicted in Figure 2. The first is an alphanumeric microfiche containing both human readable microimages and their equivalent in binary form. To create this particular format, ASCII formatted data is used. Hardcopy documents can also be stored in the form of a graphic arts microfiche. This is achieved by raster scanning each page into a digital bit stream using a document digitizer. Data compression techniques are applied to this information to reduce the data storage requirements. The third format, and the one that shows the most promise, is the mass store microfiche. This film chip has only machine readable information except for an eye-readable title block at the top. Up to 30 million bits can be recorded on one microfiche. As a way of providing a meaningful comparison, one reel of magnetic tape, with no inter-record gaps, is capable of storing 100 million bits at a cost of at least ten (10) dollars. In most real ADP applications, tape does have inter-record gaps and is seldom filled from end to end. On the other hand, a HRMR microfiche costs only thirty-five (35) cents to produce and has a large data storage capacity. In most applications one HRMR microfiche will be capable of storing the same amount of data as one reel of magnetic tape. Another advantage of this approach is that data can be rapidly accessed (in seconds) in a random manner. In contrast, retrieval speed for magnetic tape, measured in terms of minutes is severely limited by sequentially searching for the desired record.

The block diagram of the HRMR system shown in Figure 3 depicts the major subsystems of the engineering development model. The first is the recorder which operates in the following sequence. Pre-cut film is automatically transported to the laser record station at which both HR and MR data is recorded. Once this step is completed the system delivers the film chip to an on-line film processor which develops, fixes and dries the film. The next station is the verifier which will sequence through the entire digital record to ensure that all data is properly recorded. The verifier can also serve as a backup reader system. Once outputted from the verifier a metal clip is affixed, a unique identification number is assigned and the chip is loaded into the microfiche storage and retrieval module. Also included as an integral part of the automatic retrieval mechanism is another hologram reader to provide rapid data access. Controlling the entire operation of the HRMR system is a PDP11/45 computer/controller. Its function is to maintain proper sequencing of all system operations and to accomplish data transfer between various equipment modules and other systems. Data recording and readout is operated at 500,000 bits per second, which is comparable to high speed tape systems. Update of information recorded on microfiche can be rapidly and conveniently achieved. This is accomplished under computer control through the following sequence: (1) desired microfiche selected and read; (2) digital data transferred to system controller; (3) entry of appropriate data changes; and (4) initiation of record, process and verify operations. Once completed, a new, first-generation microfiche is produced that is ready for infile within the storage and retrieval module.

All information (both HR and MR) is recorded using high-speed laser scanning techniques. This method is similar to that employed by some of the more advanced Computer Output Microfiche (COM) devices currently marketed. The recorder module shown in Figure 4 consists of a low-powered, continuous-wave (CW) Argon laser. Output light is modulated by an analog signal representing the Fourier Transform of the input digital bit stream to produce the digital store holograms. Through a series of optical elements the modulated light is properly shaped, focused and scanned onto a 105 mm x 148 mm film chip. The corresponding human readable microimages are recorded in a similar fashion by deflecting the laser beam in the Y direction while the film is transported at high speed in the X direction.

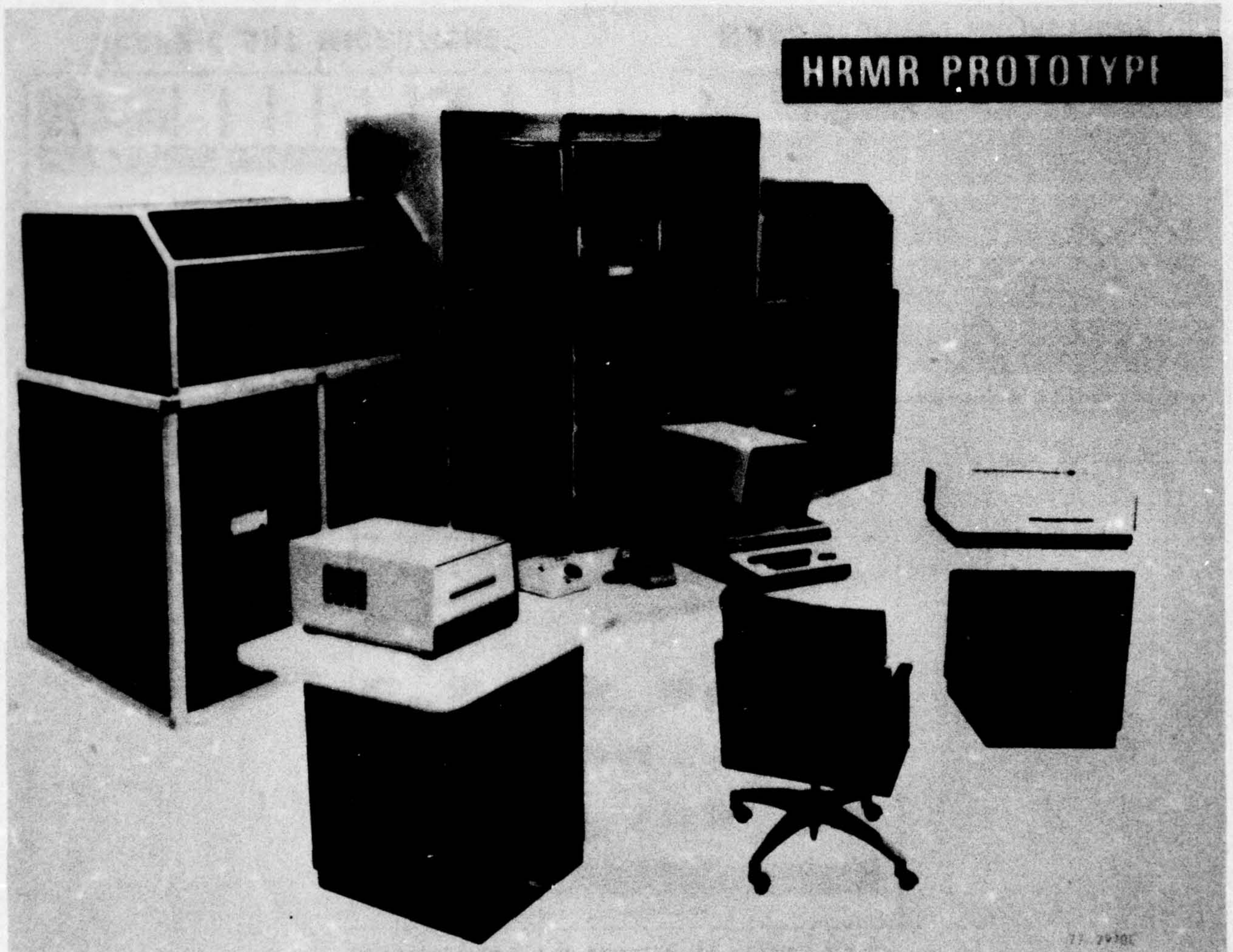


Figure 1

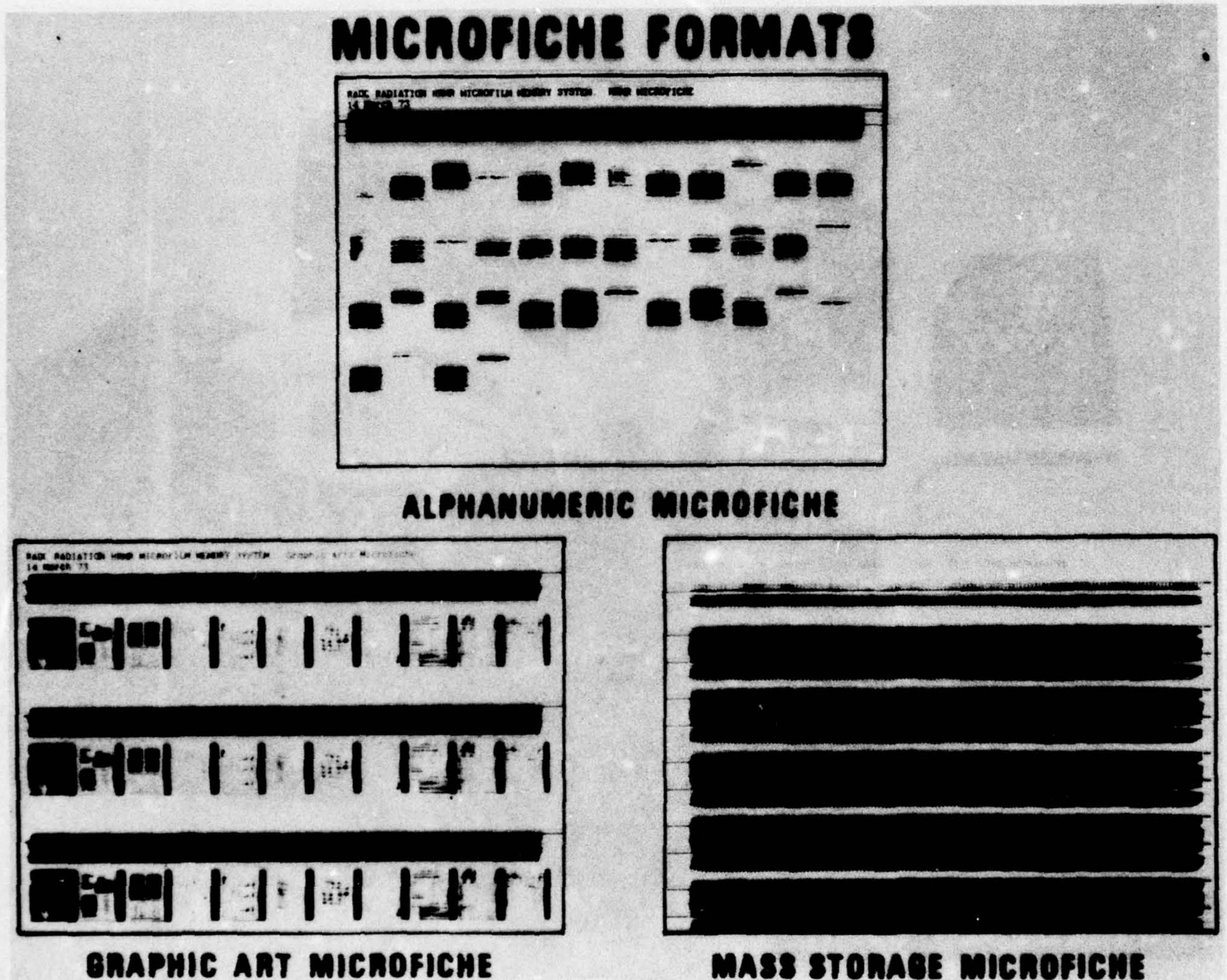

MICROFICHE FORMATS
ALPHANUMERIC MICROFICHE
GRAPHIC ART MICROFICHE
MASS STORAGE MICROFICHE

Figure 2

HRMR SYSTEM BLOCK DIAGRAM

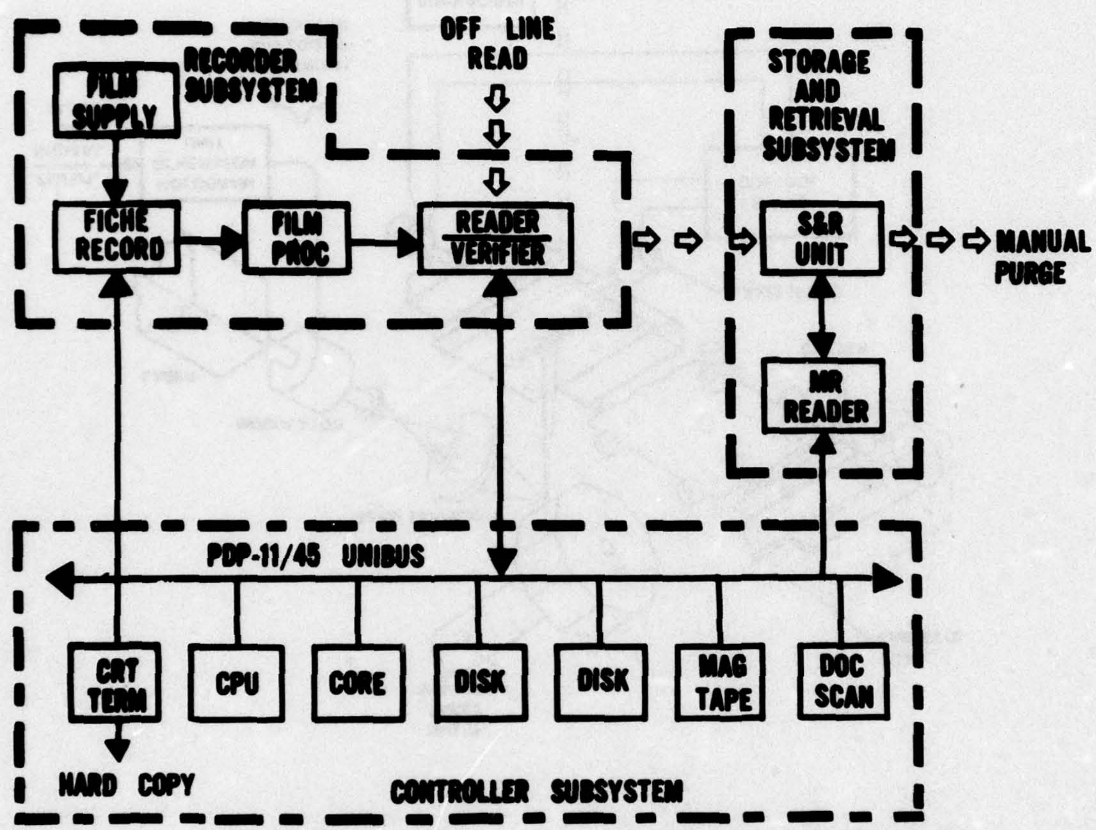


Figure 3

OPTICAL RECORDER UNIT - BLOCK DIAGRAM

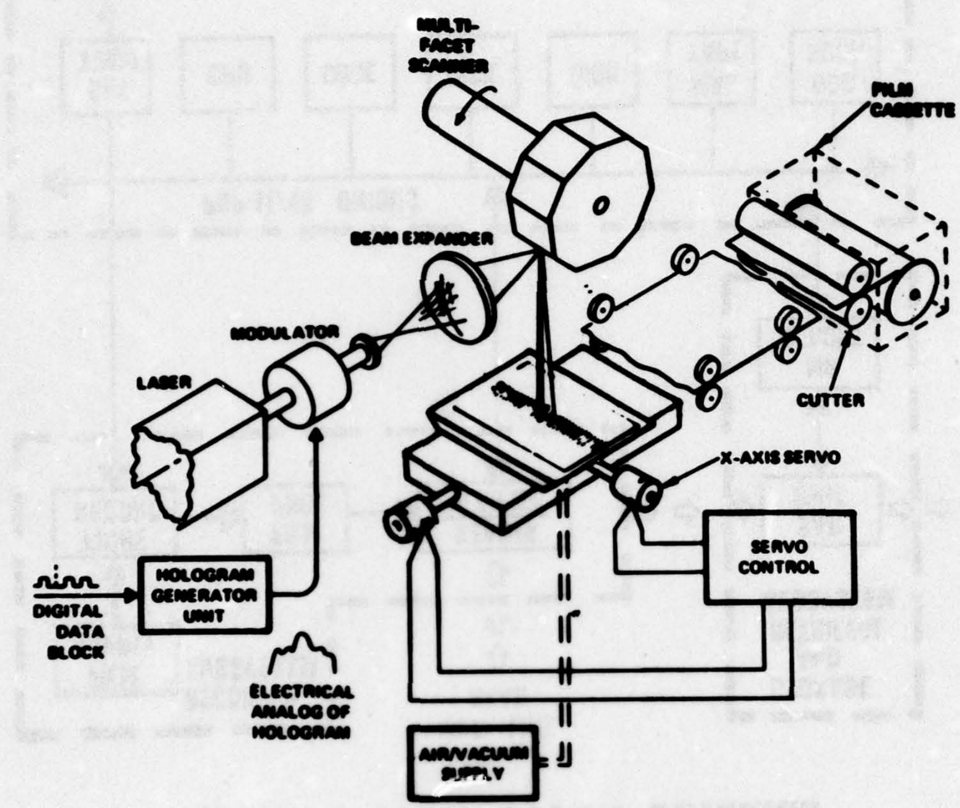


Figure 4

The basic building block of the microfiche storage and retrieval (MS&R) device is an Image System "Carousel" unit. This is an off-the-shelf item and is capable of holding 750 microfiche. Image Systems built and marketed a large number of these units for various data retrieval applications. Field performance has demonstrated their exceptional reliability and maintainability. The film extraction mechanism, supplied by Image Systems, will also be incorporated into the MS&R. As shown in Figure 5, nine carousels are stacked on a common, rotating shaft. This provides a storage capacity for 6750 microfiche. Adjacent is a hologram reader unit. This device moves in the vertical direction and can access and extract any microfiche within a maximum of 15 seconds. If the reader can imagine for a moment a device completely filled with mass store microfiche (each storing 30 million bits), a $2x10^{11}$ bit mass memory system can be readily achieved. Whenever information from a microfiche is needed by the computer, a mechanical extractor selects the desired microfiche and transports it to the read station. As user requirements increase, additional microfiche storage devices can be interfaced to the system controller via a highspeed asynchronous communication channel. This enormous storage capacity makes it possible to put the equivalent of an entire tape library on-line to a computer system.

Recovery of the digital data stored as Fourier Transform holograms is accomplished by the reader module shown in Figure 6. Low-powered, CW coherent light is used to illuminate each hologram track as it is transported at a constant rate by the film transport assembly. Light transmitted through the hologram is sensed by a linear photodetector located at the back focal plane. The data, which represents the original binary data, can now be transferred to the host computer in response to a data retrieval request.

In Figure 7, one of the Fourier Transform holograms is shown. Generation, recording and readout of these holograms is under computer control. Besides possessing very high data storage density, holograms also demonstrate data redundancy characteristics. This property makes holography highly tolerant to dust, scratches and film imperfections. Each hologram stores approximately 64 binary bits which are uniformly distributed across the entire length of the hologram. This makes each data cell relatively large (1.6 mm x 14 um), compared to the data cells required for direct binary recording (exposure of a series of spots representing ones and zeros). These large data cells are easier to locate and contribute greatly to the simplicity of the readout process. Since the data is not localized to one area but evenly distributed throughout the hologram, very low bit error rates can be achieved. As can be seen from the figure, coherent light illuminates each hologram and is reconstructed as a series of bright spots at the back focal plane. The photodetectors located at this position sense this light which corresponds directly with the original information.

To conclude discussion of the HRMR system, its key benefits are outlined in Figure 8. More information can be stored and maintained using holographic recording techniques than can be achieved using conventional magnetic techniques which allows much larger data bases to be retained for on-line computer processing. This enhanced capability will greatly improve the production of timely and complete digital products. Since all information is recorded on film, the user also realizes the significant advantage of an inexpensive storage medium capable of retaining this information as a permanent, maintenance-free record. With the dual format microfiche, both a human user and a computer can directly access desired information from a common storage medium. Finally, by employing holographic technology, low error rates and high reliablility can be achieved.

The use of holographic techniques for storage and retrieval of document-oriented material represents only one solution to the mass memory problem. This does not address the requirement for the storage and retrieval of digital data in extremely long records and at very high recording and readout rates, such as those presented by a satellite transmission. Typical requirements extend into the 1.0 to 2.0 gigabit per second (GBPS) range with error rates not to exceed 1 bit in 10^6 bits required. Investigation into the more "well established" wideband recording techniques, i.e., magnetic longitudinal and rotary head, laser and electron beam and their predicted capabilities into the 1980 time-frame yield results similar to those shown in Figure 9. This rapid and continuing growth in data rates is far beyond the capabilities of current data handling equipment. As illustrated, the more conventional recording schemes cannot approach the 1.0 to 2.0 GBPS rates under consideration. To meet these demands, Rome Air Development Center (RADC) under a cost-shared Air Force Systems Command, Directorate of Laboratories and Advanced Research Projects Agency (AFSC (DL)/ARPA) program has demonstrated a 750 megabit per second (MBPS) recording capability with possible extension to 2.0 gigabit per second (GBPS). The concept employs a holographic rather than bit-by-bit laser recording scheme.

MICROFICHE STORAGE & RETRIEVAL SUBSYSTEM

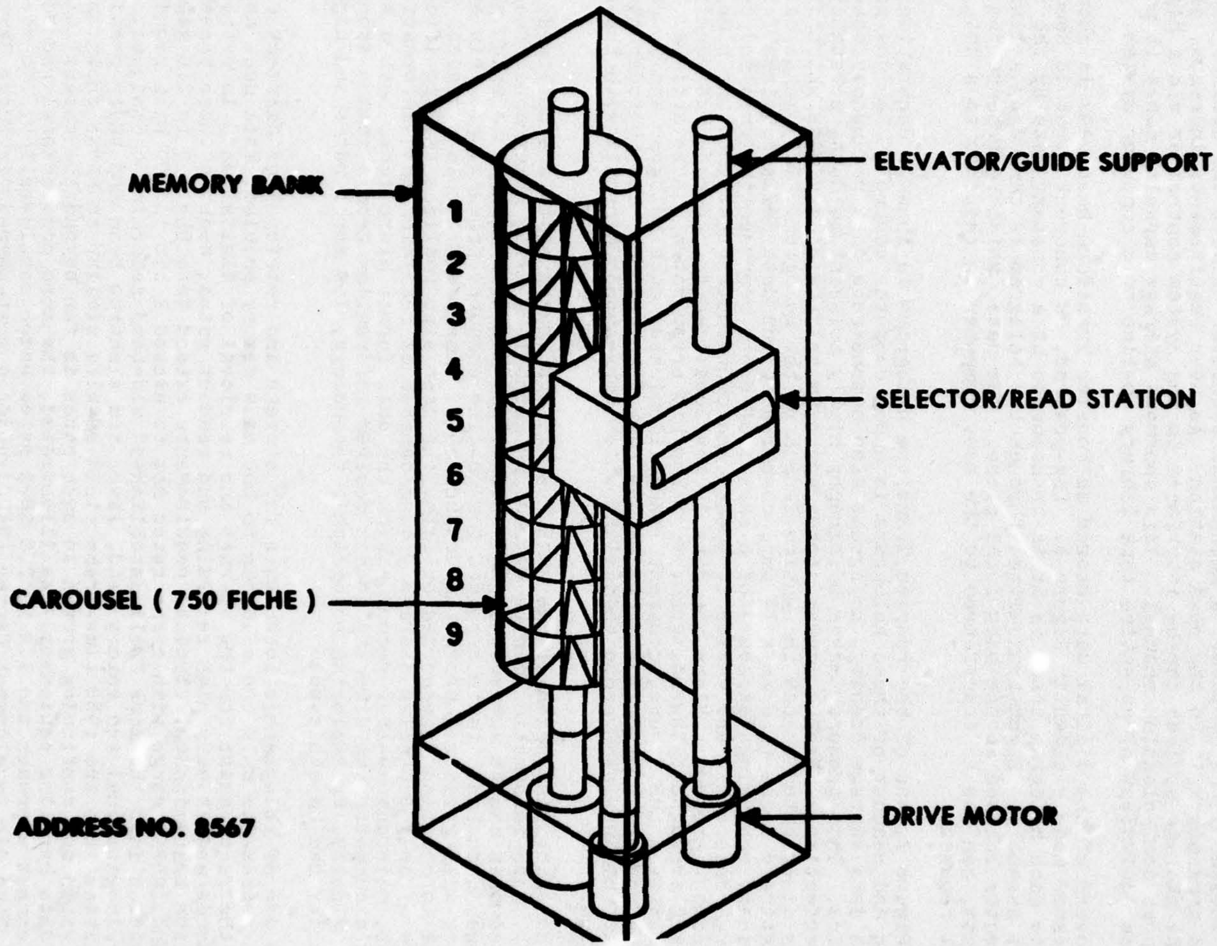


Figure 5

DIGITAL DATA READOUT MODULE - CONCEPTUAL DIAGRAM

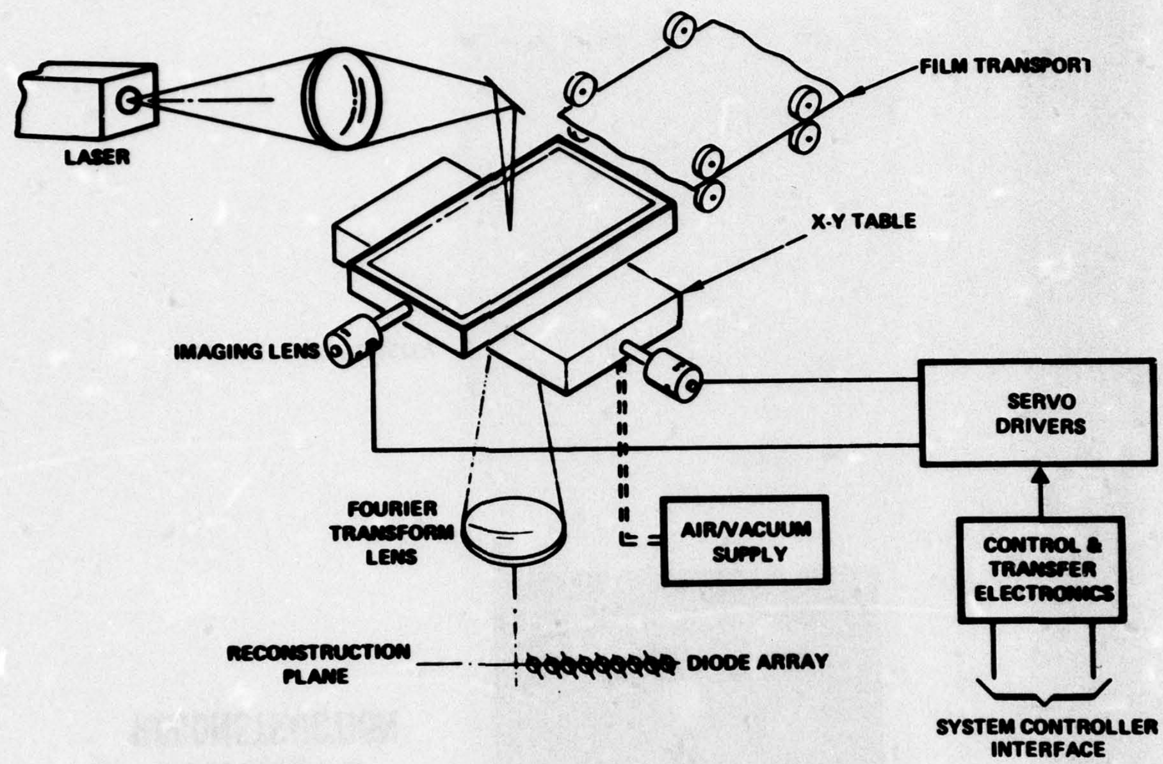
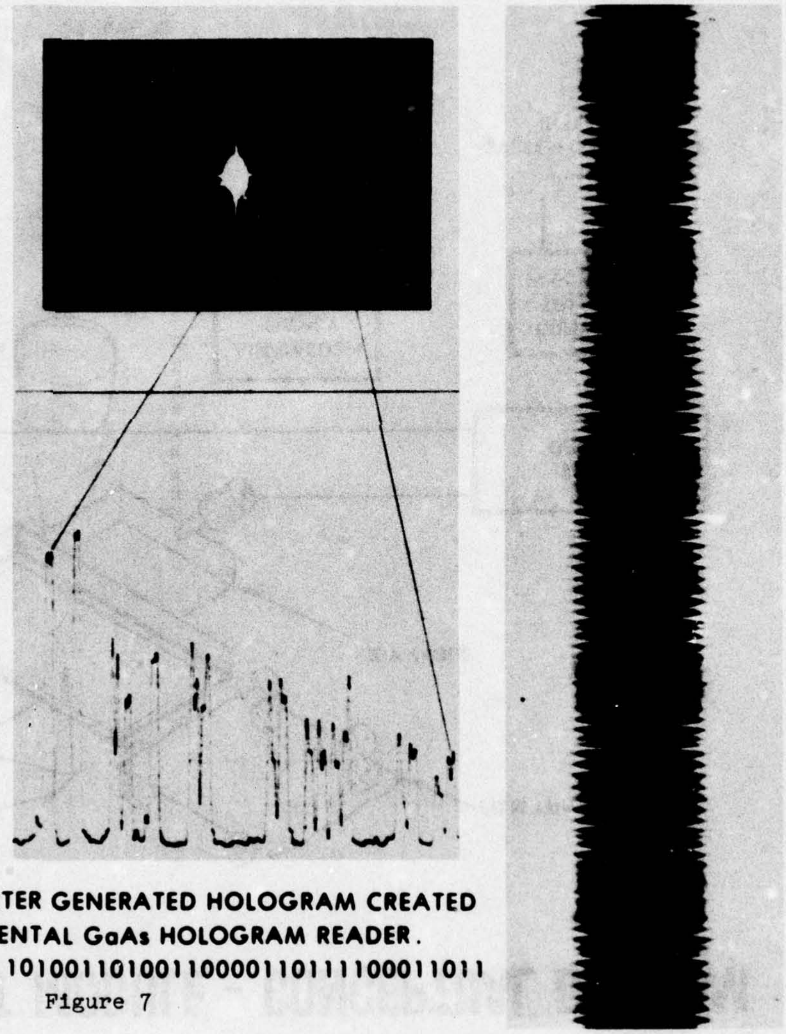


Figure 6

COMPUTER-GENERATED HOLOGRAM & RECONSTRUCTION

INTENSITY



RECONSTRUCTION OF COMPUTER GENERATED HOLOGRAM CREATED FOR READOUT WITH EXPERIMENTAL GaAs HOLOGRAM READER. THE BIT SEQUENCE SHOWN IS 10100110100110000110111100011011

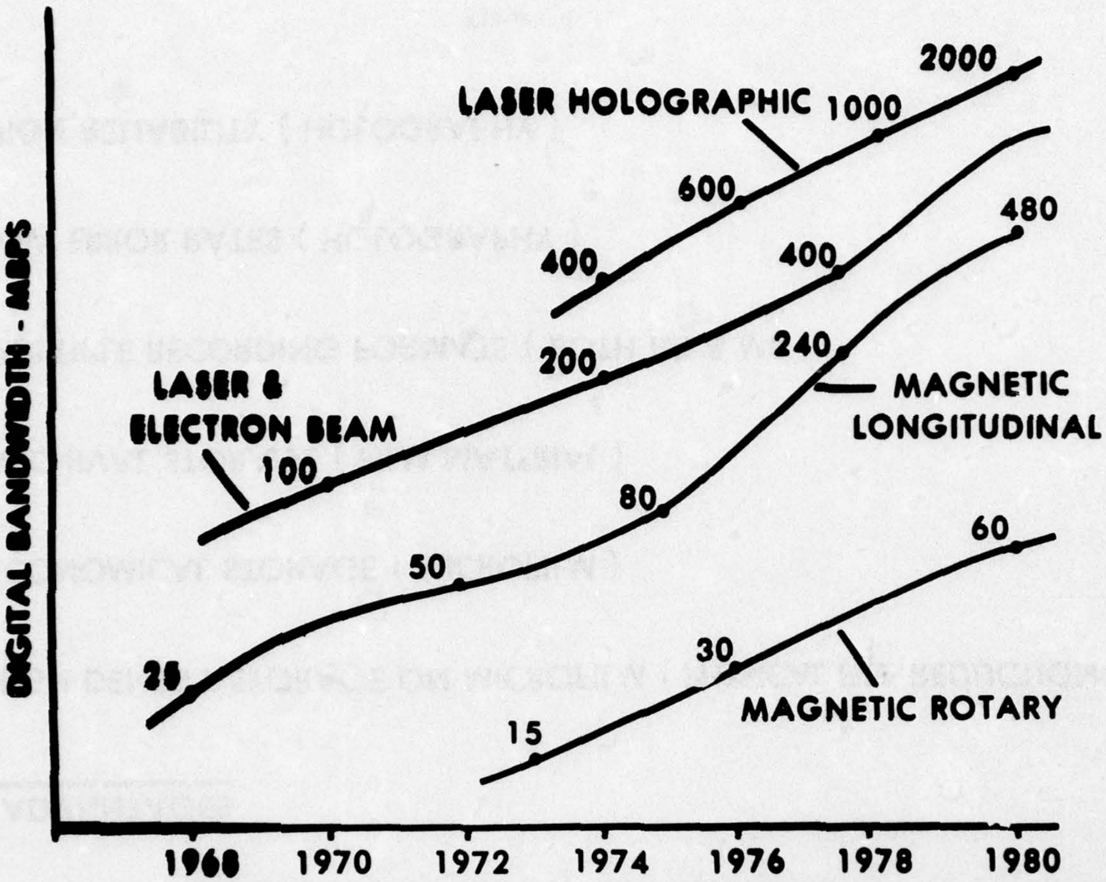
Figure 7

HOLOGRAPHIC MICROFILM MASS MEMORY SYSTEM

ADVANTAGES

- HIGH DENSITY STORAGE ON MICROFILM (PHYSICAL FILE REDUCTION)
- ECONOMICAL STORAGE (MICROFILM)
- ARCHIVAL STORAGE (FILM MATERIAL)
- MULTIPLE RECORDING FORMATS (BOTH HR & MR)
- LOW ERROR RATES (HOLOGRAPHY)
- HIGH RELIABILITY (HOLOGRAPHY)

Figure 8



WIDEBAND RECORDING - PREDICTED PERFORMANCE

Figure 9

As shown in Figure 10, the recording medium is 35 millimeter (mm), silver halide, photographic film in roll form. Incoming serial data is demultiplexed into 128 parallel channels. One bit from each of the 128 channels is stored in parallel in a one-dimensional hologram. Holograms are recorded across the film width in rows and measure 0.8 mm in length on 0.016 mm centers. In addition, there is a 0.2 mm guard-band between hologram rows. The active recording area of the film is 24.2 mm, with the system scanning 1512 holograms per row. This results in an average packing density equivalent to greater than 800 thousand bits per square centimeter.

Figure 11 depicts a simplified configuration of the Wideband Holographic Recorder. The diagram shows the externally modulated continuous wave (CW) Argon laser used in both recording and readout. The beam-forming optics split the laser beam into both a signal and a reference beam. Both beams pass through the acousto-optic page composer, which generates a 128 bit spatial pattern and a 1 bit reference beam. The faceted mirror scanner sweeps both patterns across the width of the 35 mm film through a transform lens. The Transform lens generates the Fourier Transform hologram of the data. This results in the information contained in the 128 data bits being distributed throughout the entire, relatively large, area of the hologram.

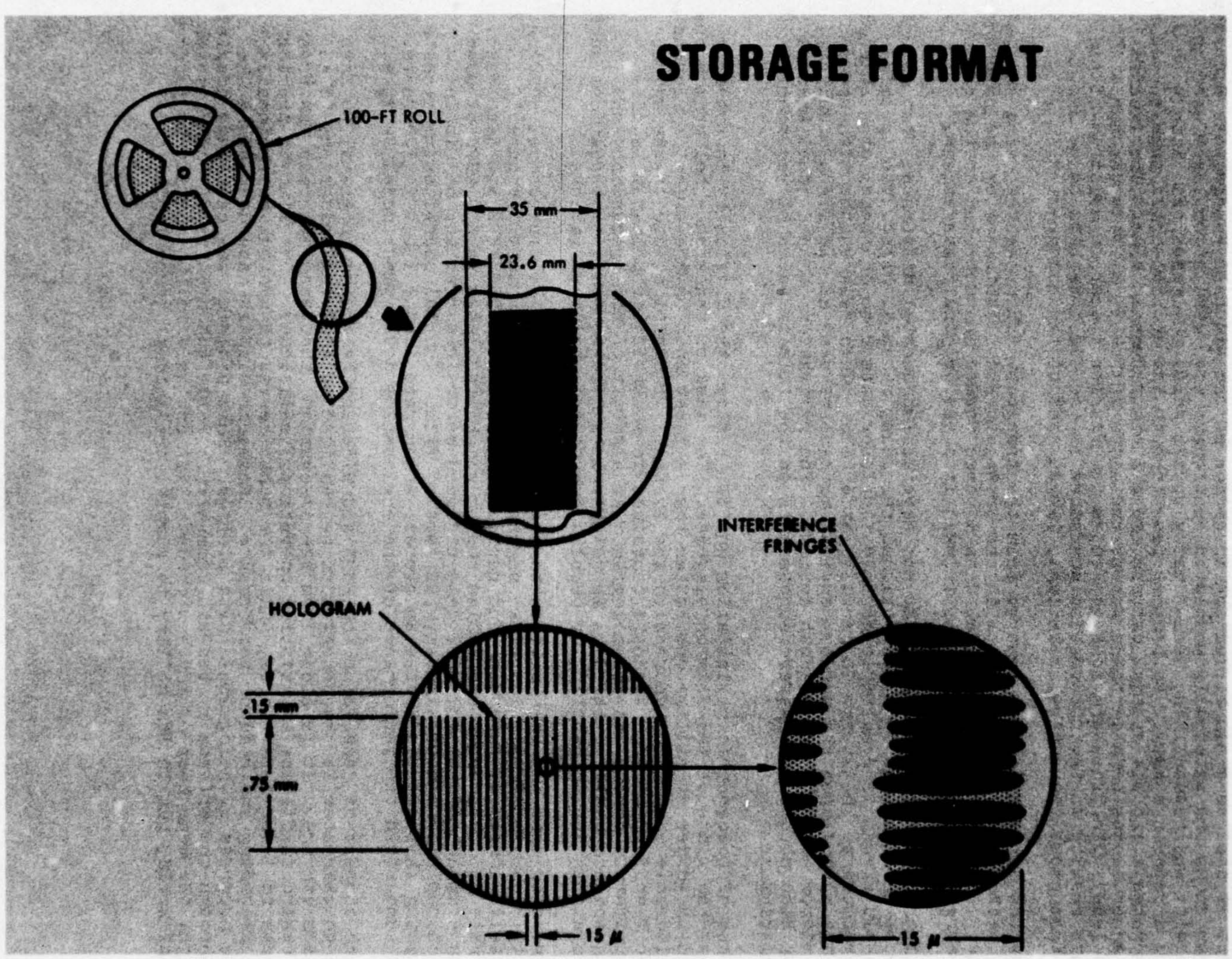
After the exposed film is removed from the transport and processed using conventional techniques, it is replaced in the transport and illuminated by the reference beam alone. An image of the original data is thus formed and detected at the photodetector array. The data is then multiplexed into the original data stream.

A holographic recorder/reproducer has several fundamental diferences that have proven to be advantageous over conventional recording schemes. First, the incoming high rate bit stream is demultiplexed into 128 lower data rate channels, each channel being introduced to the system in parallel at 6 MBPS. This relaxes the high-speed requirements placed upon the various optical mechanisms such as the light modulator, spinning mirror assembly and readout devices. The relaxation of individual channel bandwidths also places less severe constraints on the film in terms of reciprocity characteristics.

The use of holographic registration on film reduces the difficulty in tracking data across the width of the film on playback. The task of tracking a hologram whose area is on-the-order of millimeters rather than a laser spot on-the-order of microns is a considerable advantage inherent to the system. In addition, due to the nature of Fourier Transform coding, the data contained in a particular hologram is distributed uniformly throughout the hologram. As a result of this redundancy and spatial invariance quality, the data is relatively immune to media imperfections such as film scratches due to handling or bubbles due to processing variations.

Progress to date has been satisfactory and reported results impressive. During January, 1976, an implementation of an engineering development model (Figure 12) proved itself capable of recording and reproducing data at a 750 MBPS raw data rate. This raw rate includes error correction coding which translates to 600 MBPS of user data. The playback of the 750 MBPS data produced error-free readouts on single selected channels over approximately 200 feet of film. Initial quantitative results from analyzing this data indicates an overall error rate in the 1 bit in 10^6 range. Figure 13 summarized the performance of the engineering development model to date.

In the context of describing the HRMR and Wideband Recording programs, we have attempted to characterize the benefits to be realized by applying holography to digital data storage. It is clear from the discussion, that each approach has been directed toward fulfilling an existing operational deficiency. One aims to provide ready access to large data stores, while the other serves to provide exceptionally high data recording and readout performance. Research studies have documented the impressive progress of holographic data storage. This progress has been made possible largely through the efforts of the R&D community. While the promise of these system remain high, the true test of any new product lies not only in its ability to perform in the real world, but in its ability to surpass traditional technologies.



86665-13A

Figure 10

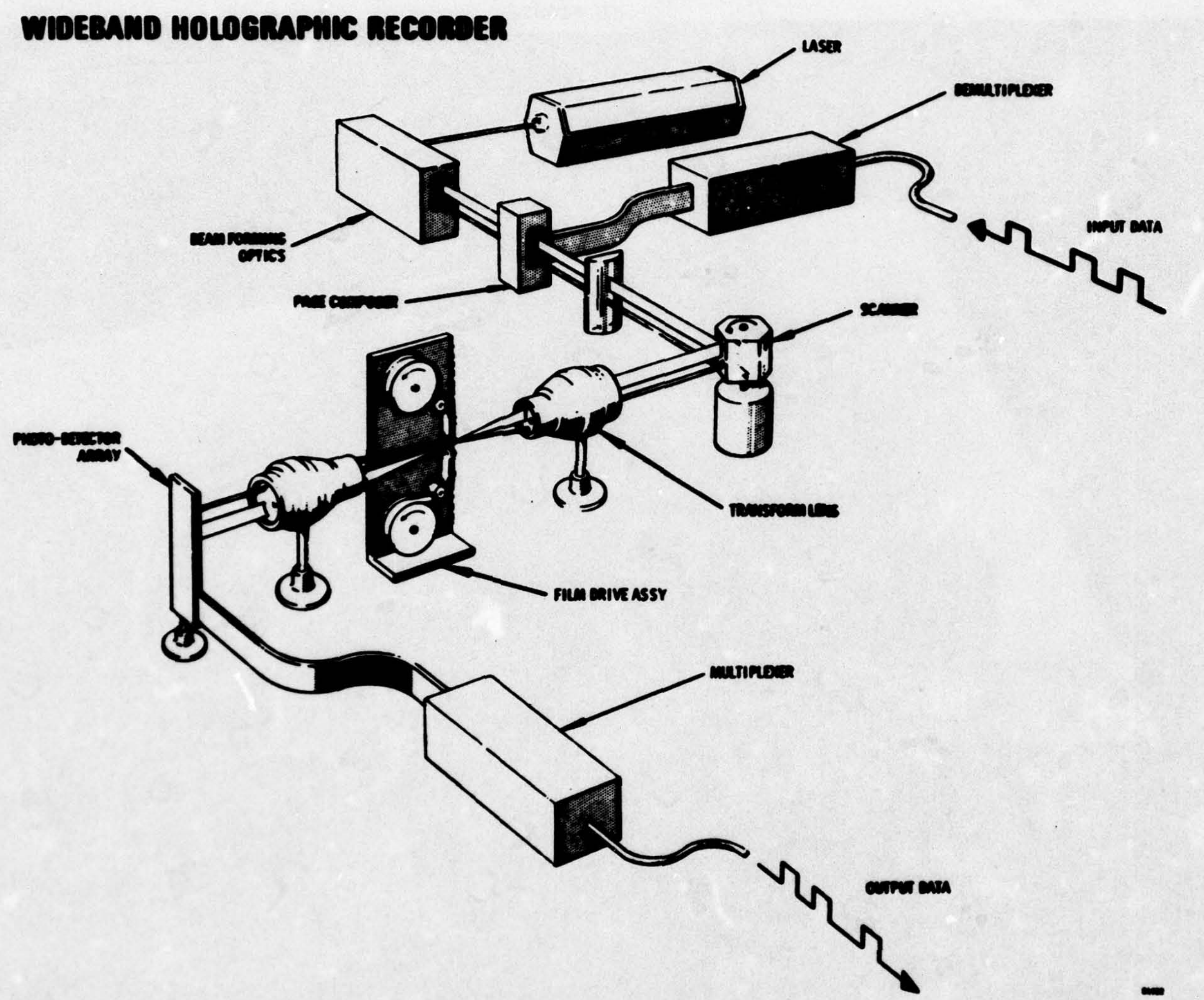

WIDEBAND HOLOGRAPHIC RECORDER
LASER
DEMULTIPLEXER
INPUT DATA
BEAM FORMING
OPTICS
PAGE COMPOSER
SCANNER
PHOTO-DETECTOR
ARRAY
TRANSFORM LENS
FILM DRIVE ASSY
MULTIPLEXER
OUTPUT DATA

Figure 11

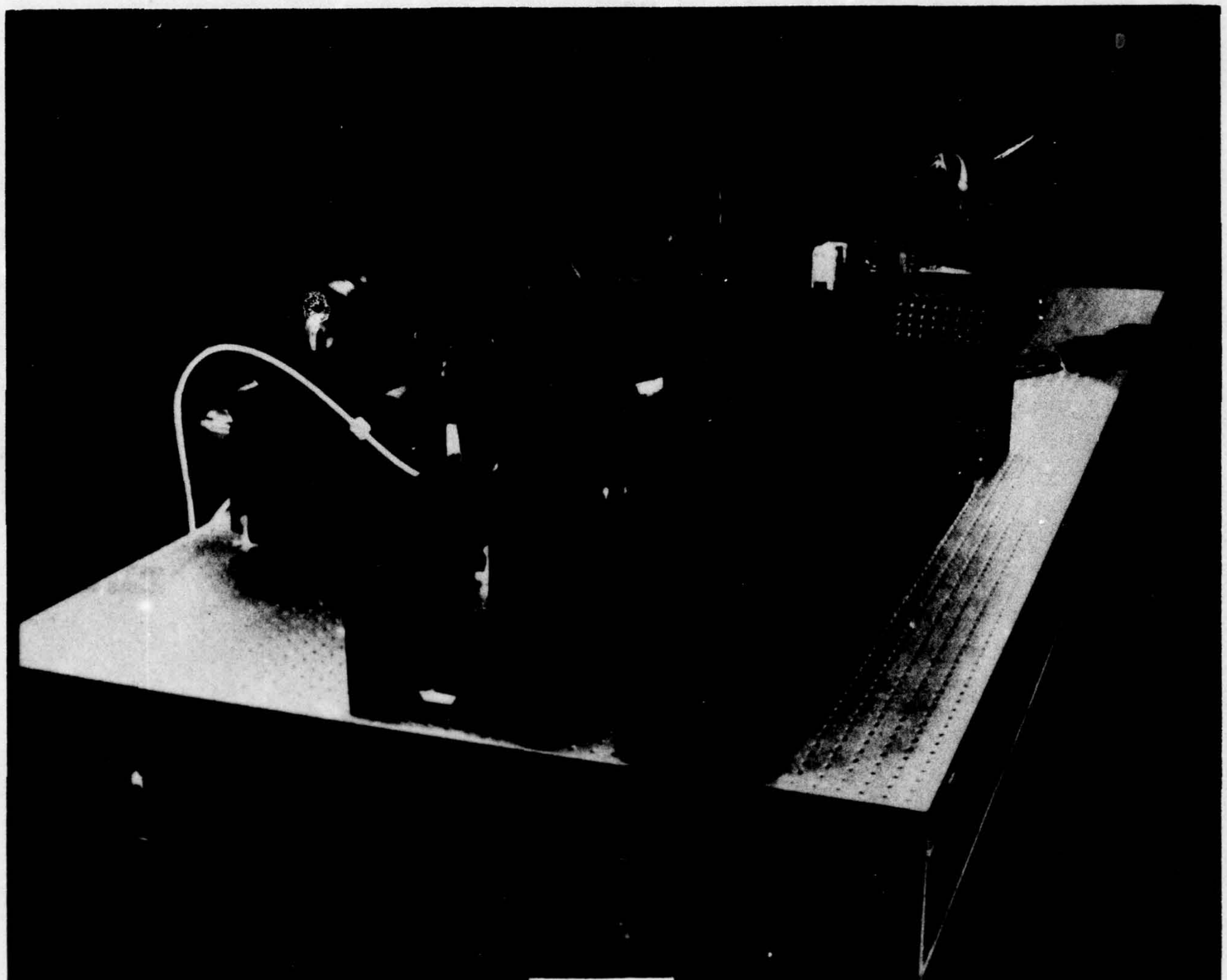


Figure 12

PERFORMANCE RESULTS OF THE WIDEBAND HOLOGRAPHIC DIGITAL RECORDER

RECORDING

TOTAL DATA RATE	750 MBPS
CHANNEL DATA RATE	6 MBPS
NUMBER OF CHANNELS	128
FILM VELOCITY	3.8 m/sec
PACKING DENSITY	0.8×10^6 bits/CM^2
HOLOGRAM EXPOSURE	3 mw x 90 nsec

READOUT

TOTAL DATA RATE	750 MBPS
CHANNEL DATA RATE	6 MBPS
SIGNAL-TO-NOISE RATIO	18 db
BIT ERROR RATE	10^{-5}

Figure 13

MULTIMODE NETTING BY WIDEBAND CABLE

VICTOR A. DEMARINES and GEORGE A. FAGAN
The MITRE Corporation
Bedford, Massachusetts, U.S.A. 01730

SUMMARY

The rapid growth of low cost digital circuitry is affecting all forms of electronic technology. Because cheap and reliable devices are becoming more and more available, the concept of information processing is altering. Digital processing is still important, but is being augmented by analog devices - video and audio recordings, microfiche, etc. Digital processing itself is altering, as distributed processors and smart terminals become common. Conventional communications technology does not adapt well to the new demands. The multimedia communications bus, based upon current television coaxial cable technology, offers an economical, reliable and flexible alternative.

INTRODUCTION

It is hardly necessary to state that the field of information processing is changing - that seems to have been true for the last twenty years - and may well be true for the next fifty. Many of the changes are due simply to developing technology, but beyond that, there are changes in concept, changes in the way people think of information systems, that will have significant effects upon the architecture of systems in the future. Because military organizations have been, and probably will continue to be, leaders in systems development and application, they are among the first to feel the impact of the changes.

In the first place, the concept of information processing has broadened considerably. Often, in the past, the term was used almost as a synonym for digital data processing, with a possible extension to include on-line retrieval and manipulation. The link with the digital computer is still there, but non-digital forms are also being recognized as useful and necessary parts of an information system. We now find the various forms of voice communications - telephone, radio, secure or open, being considered as part of an integrated design rather than being accepted as separate and independent systems. Television also is starting to assume a significant role, not merely for entertainment, but in such varied applications as surveillance, teleconferencing, remote access to documents, and interactive education systems. Digital computer systems and digital communications are, of course, still important, but they are no longer the sole concern of the information system designer.

Secondly, the development, during the last few years, of low cost digital devices is changing the approach to computing. Computing power is no longer rare and expensive and the designer is no longer forced to think in terms of maximizing the utilization of an expensive central computer facility. Mini-computers and smart (in some cases, brilliant) terminals can be dispersed throughout a facility for individual use. The distribution of processing and the resulting need for remote access to data bases that may themselves be physically dispersed creates new problems of communication. A great deal of effort is therefore being put into the development of effective computer netting systems, and though these may reduce the inefficiencies and delays which have occurred in large time-sharing systems, they place additional strain on the communications system, because large portions of the data base may be shared by many devices. The volume of digital traffic between dispersed computing facilities and a logically central data base may be considerably greater than that between dispersed terminals and central computing facilities.

One of the most revolutionary effects of the ready availability of new technology is that the conventional paper systems, on which most of the world's organizations depend, can now be seen to be wasteful of manpower, materials and space. In the past few years the same pattern has emerged in most, if not all, of the industrialized countries; mail services have become less efficient, less timely, and more expensive, while the scope of services has been reduced. In the case of national postal services, the results are clear for all to see; similar problems occur, but are not so visible, in the internal operations of large industrial companies and military organizations. Various forms of electronic mail systems are therefore being considered and evaluated. These include for example digital message transmission, facsimile, video, remotely accessed micro- and ultra-fiche, etc. The range of possibilities is great, but, in any organization where the volume of internal mail is sufficient to justify an electronic distribution system, the required number of terminals and the volume of electronic traffic would be very large.

Historically, and conventionally, communications systems have been developed and added to the existing base - telegraph was added to mail, telephone to telegraph, then radio, then television, then digital transmission. Each different type of service tends to have its own physical plant and support organization and generally one of the ground rules for adding a new service is that existing services

must not be modified or impaired in any way that is visible to the user. This approach, though unavoidable, has led to the creation of multiple services, sometimes redundant, sometimes wasteful of resources, and often so inflexible as to inhibit and obstruct the introduction of new features. The multiplicity of separate services, which are competitive rather than cooperative, is not compatible with the integrated information system; a comprehensive and flexible system will be needed to support the emerging applications. This system must provide the ability to transmit information in a variety of forms, among a large number of users, in timely fashion.

The concept that is being used and developed at MITRE draws upon two technologies, cable communications and low cost digital electronics, to support multimode "bus" systems. The cable is a conventional coaxial cable as is used for closed circuit or community antenna television; the modes include voice, video, audio, and digital transmissions; the terminal devices include telephones, television cameras and monitors, teletypes, cathode ray displays, computers, printers, and entry control devices all attached to and communicating through a single coaxial cable distribution system. The term "bus" implies that all information is available to all participants. The advantages of such a system are:

flexibility	- new services and new subscribers can be added easily,
cost	- not only is the installation cost much lower than that of the conventional separate systems, but the interfacing devices are also inexpensive.

At present MITRE is involved in experiment, application and system design using this type of multimode communications for a number of Department of Defense and other sponsors. We are developing the concepts and devices for application to Army Base systems, large government agencies, Air Force Command Centers, and hospitals, but we can foresee their utilization in a great variety of communities where there are significant problems in maintaining adequate communications. Military and governmental organizations have such problems, but they are not alone; industrial corporations, financial institutions, manufacturing plants, and universities have similar problems and could use similar solutions.

The pressures created by the rapid growth in electronic, and particularly digital, technology are forcing a reappraisal of the conventional communications structure. Cable-contained bus communication is a vehicle that is needed to support new applications economically in a wide variety of closed communities, and we believe that there will be tremendous increase in its use in the next few years.

CABLE TECHNOLOGY

The notion of coaxial cable as a multimode transmission medium tends to be distorted by the underlying association with CATV - Community Antenna Television - and the feeling that the cable is somehow "full" of television channels. To dispel that idea we must consider the cable simply as a radio frequency medium. It has a frequency dependent attentuation characteristic but despite this, common commercial practice provides a usable bandwidth up to 300 megahertz, and in some experimental applications, as much as 900 MHz has been used. As the cable is shielded, it is virtually free from interference. Therefore the cable provides a capability that can be employed in a wide variety of ways just as radio communication can be employed. Frequency bands can be allotted to specific purposes and within those bands well known modulation and multiplexing techniques can be employed.

There is nothing strange or novel about the use of wideband cable. The research and development was done long ago, and the devices that are required to support wideband transmission over long cable lengths are available from stock from a number of suppliers. The cable television industry maintains a standard 80 dB isolation for all components of the distribution network. This means that, in exceptionally noisy environments, such as manufacturing plants with much heavy equipment generating electromagnetic interference, the noise level in cable is low enough to provide excellent performance. The practice in the industry is to maintain an adequate signal level by installing frequency compensated linear repeaters at distances along the cable corresponding to 22 dB attenuation at the highest frequency of interest; typically, a repeater, which costs about $300, is installed every 1000 feet (300 metres). Cable itself costs about 10 cents per foot (33 cents per metre), and installation cost, when the cable and repeaters are hung on existing telephone poles, is about four times the cable cost. An average figure for the total installed cost of a cable and repeaters is therefore about 80 cents per foot ($2.65 per metre), but this will of course vary from one installation to another. For two way transmission, it is convenient, but not essential, to install a pair of cables; because of the installation cost, which is not doubled if two cables are hung at one time, such an arrangement would cost about $1.30 per foot ($4.30 per metre).

The technique for adding subscribers to a cable is also well established - simple cable taps are used and all subscribers potentially have access to all the information on the cable; actually, access to services is limited by the type of interfacing equipment supplied to the subscribers, just as is the case for broadcast radio. Controls are obviously needed to prevent unauthorized access which can range from theft of a service (e.g. television, or computer time) to disclosure of classified data. If the potential risks are recognized during the design phase, protection can be built into the system; the real dangers are therefore less severe than they at first appear.

The costs and the performance of this type of cable installation have been well established by several years of practical field experience of cable television in many cities in both the United States and Europe. Costs are low, reliability is high, performance is excellent. The availability of the radio frequency spectrum is illustrated by Figure 1, which shows the signal spectral pattern in use on a multi-mode cable at MITRE. This installation is providing television (both commercial and closed circuit), frequency modulated commercial radio, voice telephone, and digital services, simultaneously.

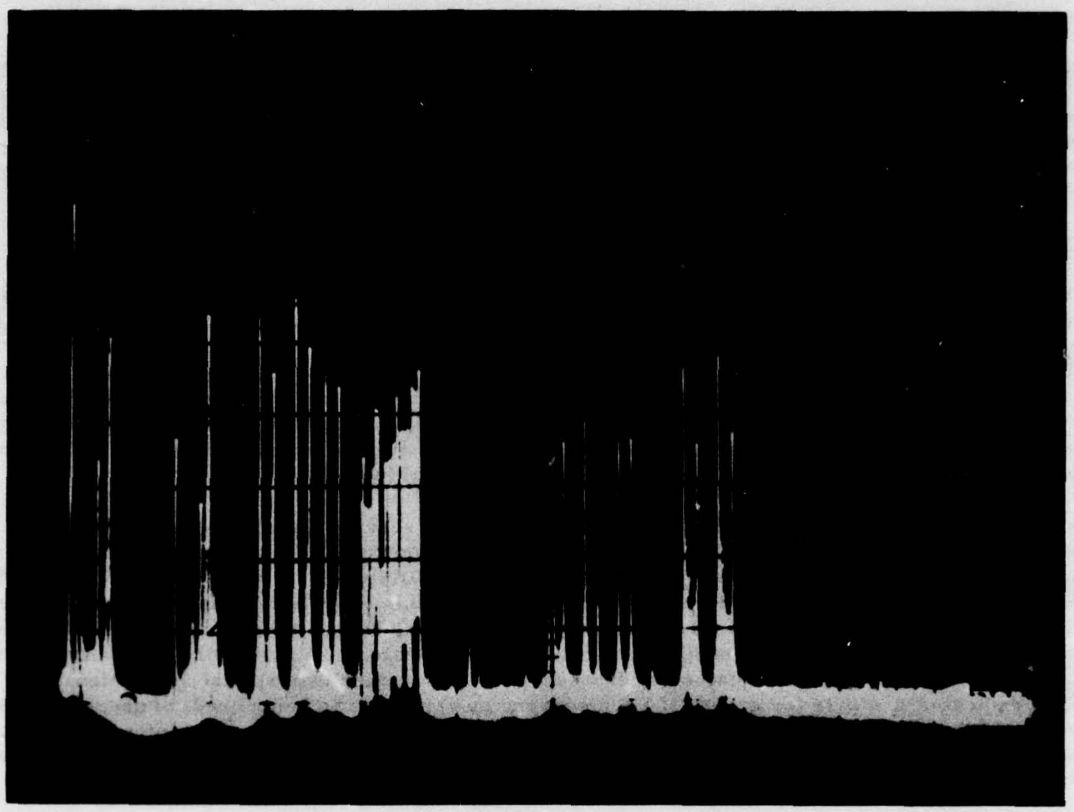


Figure 1. Frequency Spectrum of the MITRE Demonstration System Cable

The provision of cable plant in a community not only will support a wide variety of services, but also makes it possible to use a single control facility for fault detection and isolation. Also it makes it possible for any user to connect to the information system, anywhere in the community at little more than a moment's notice, for only a simple tap is needed. In the remainder of this paper we propose to discuss briefly the conventional services, and to dwell in some detail on the digital data communications system in use at MITRE.

CABLE TELEVISION SERVICES

By far the most commonly known use of coaxial CATV cable is, of course, for the transmission of television. Such services, whether for off-the-air television, closed circuit television (e.g. for conferencing or surveillance systems) and even for many computer assisted instruction applications, can be accommodated in 4.5 megahertz bandwidth within 6 MHz channels. Twenty such channels could therefore be available on the cable and use only 120 MHz of the available 300 MHz bandwidth - obviously there is room for other services.

Some television-like services require higher resolution and therefore more bandwidth than conventional television. For instance, the multicharacter video display terminals that are used in some computer aided instruction applications may require 15 MHz channels. Dissemination of microfiche for remote viewing requires extremely high resolution to accommodate the fine print and line widths of many documents, and could require 100 MHz or more for continuous transmission. Fortunately, these types of application rarely need continuous transmission. The viewer normally spends a considerable time, perhaps several minutes, studying a single image so the strain on the transmission can be reduced by providing local storage devices (so-called frame grabbers) from which the image can be reproduced at the remote display terminal, in which case the high resolution bandwidth can be dedicated to a user for only one frame tranmission time of about 30 milliseconds, to fill the storage memory. In addition, if slow scan techniques are used, the bandwidth requirement can be reduced, though the

frame transmission time would be increased proportionately, and therefore the number of users who could be serviced on a single channel in a given interval would also be reduced.

Obviously this kind of system offers the designer room for compromise. He can dedicate the bandwidth to a single user or to multiple users who want to view the same frame; he can frequency divide the band into several narrower channels and serve a number of users either with less resolution or with slower response; he can time divide the band, or the separate channels, and, if necessary, address selected single frames to individual storage devices.

The various techniques place different demands upon the signal transmission and reception equipment, but any and all can be supported by a single coaxial cable. The selection of the appropriate combination of techniques to satisfy a given organizational need can be based upon well understood tradeoffs between cost and capability.

VOICE TELEPHONE SERVICES

In the conventional telephone system, each subscriber has a pair of wires that run from his instrument to a control exchange where switched connections are made in response to dial signals. In all but the most modern exchanges, switching is a mechanical operation, and the arrangement is relatively inflexible; if a subscriber changes his location the old wires must be physically disconnected and the new ones connected to the appropriate terminals in the exchange. The number of subscribers who can be supported is limited to the numerical capacity of the exchange equipment. Additions are expensive and time consuming, and yet, because most telephones are used for at most a few hours per day, most of the equipment and lines are idle for the greater part of their working lives.

A single coaxial cable can be used as the only signal path to service a telephone system. In one operating commercial system, up to 3000 telephones can be supported; more are obviously possible if bandwidth is allocated. In this approach, frequency division multiplexing is used to connect different subscribers. Switching is performed by frequency selection within the telephone instruments, under the control of a central, stored program minicomputer that allocates, by digital switching, the calling and called instruments to any vacant channel. Each channel is the analog of a wire pair, but communications privacy and non-interference are obtained by frequency division rather than spatial separation. A channel is dedicated to a pair (or group, if the conferencing capability is being used) of users only for the duration of the call; it is then available for use by another pair, so the common equipment can have a much higher duty cycle, thus giving a better return on the investment than conventional switched wire systems.

Adding new subscribers to the system simply requires taps into the cable, and a variety of useful services can be provided by programming in the minicomputer.

DIGITAL COMMUNICATIONS

It is fairly obvious that, if the bandwidth of the cable can be separated into discrete channels and modulated to carry video or audio signals, then digital messages can also be carried by frequency shift keying and similar methods. Conventional techniques, however, have the disadvantage that, if a very high data rate channel is required for a short time, a disproportionate amount of bandwidth must be allocated to it - that is, other services may be impaired because of the need to keep a fixed capacity available, even though it may be used infrequently, and the user pays either directly or indirectly for the capacity available rather than the service used.

At MITRE an attempt has been made to solve some of the problems by developing a flexible system in which the available capacity can be reallocated according to current needs. This is a transparent high-speed burst transmission system that is designed to provide fast access by many users to variable data-rate digital channels. The system could operate over any wideband medium, but in the demonstration system a coaxial cable is used.

MITRIX DEMONSTRATION SYSTEM

MITRIX is a high-speed burst transmission system operating on a pair of cables - inbound and outbound (Figure 2). In the demonstration system at MITRE, the digital communications system occupies only a very small fraction of the cable bandwidth - about one MHz, but this is sufficient to support an aggregate data transmission rate exceeding 600 Kbps.

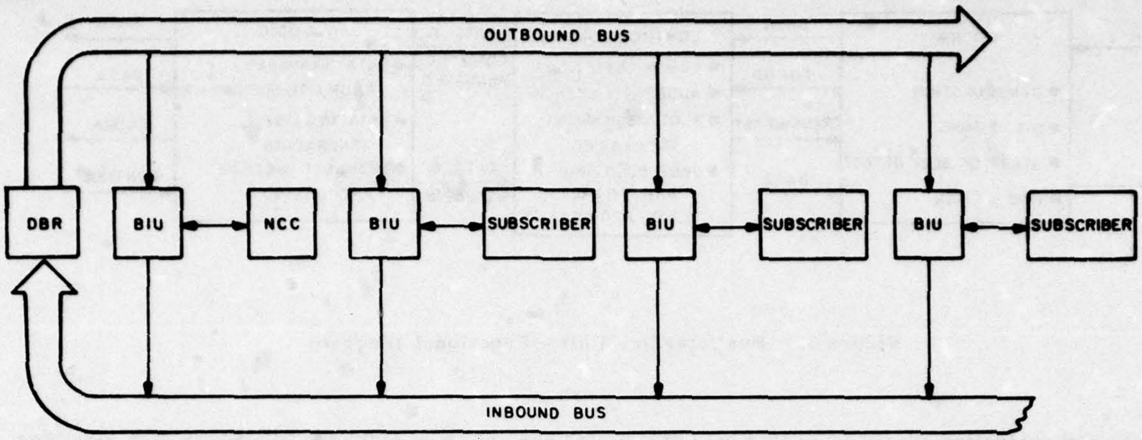


Figure 2. MITRIX Functional Configuration

The channel carries a serial bit stream that is divided into frames of fixed time (and data) length. Each of these frames is divided into a sequence of fixed length data slots. The system operates as a time division system in which a user is allocated certain data slots in each frame for transmission. Allocation of the slots is controlled by a network control center according to the data rate needs of the sender. For example in the demonstration system, an allocation of one slot per frame (i.e. a 256 bit burst containing 192 information bits in each 2.56 second frame) corresponds to a data rate of 75 bps. Slots may be allocated in multiples of 2^N, thus permitting communication at 150, 300, etc. to a maximum of 307,200 bps. The allocations remain in effect only until the end of a transmission, when the slots are released and may be reallocated to a variety of other senders.

As this is a bus system, every user receives all data slots. Selection is performed by the receiving equipment which uses part of the data in each slot, the address, to identify the slots intended for it. The address can designate a single user, thus effectively giving point-to-point communication, but may also indicate a class of users (including "all") and thus it is possible to distribute data in conference or broadcast modes or to poll a group of users. As a receiver selects slots of interest from all on the bus, it is obviously possible that the receiver may receive and process several messages at different data rates, simultaneously.

The interface between the user and the cable is the bus interface unit (Figure 3), a rather sophisticated device, but one that, because of the availability of low cost digital integrated devices, need not be expensive. The bus interface units are connected at each user's location, between the inbound and the outbound cable, and each is also connected to one or more terminal devices. The unit accepts information from an input device, reorganizes the data into suitably addressed packages, and inserts them into the inbound cable in the assigned time slots. The serial bit stream formed by all the active bus interface units passes through a digital bus repeater on to the outbound cable, from which individual packages are selected, according to address, by other bus interface units. The address information is then removed and the data that was originated by the input device is reconstituted and passed to the receiving device.

The Network Control Center is a special terminal device which performs three principal functions: system initialization and start up, changes to slot allocations to accommodate different data rates, and monitoring and fault isolation. The Network Control Center can be programmed to detect terminal device failures and to reallocate some of the functions of a failed unit to a substitute device. Therefore, in systems in which there is duplication of equipment, the system can be protected against catastrophic failure by reconfiguring the network for continued, even though degraded, operation.

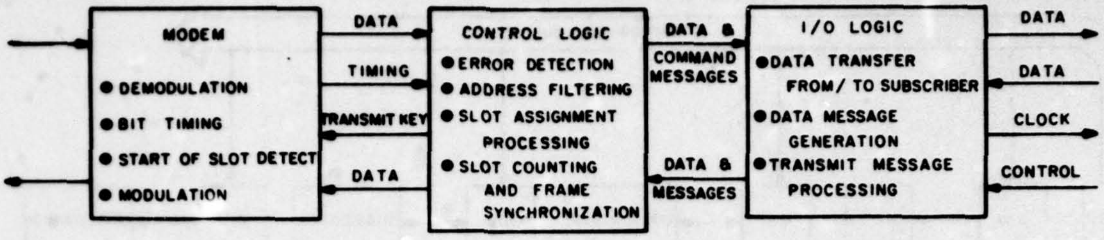


Figure 3. Bus Interface Unit - Functional Diagram

The demonstration system operates only one digital bus, at about 600 Kbps aggregate data rate, but even this can accommodate 8000 users at the minimum data rate of 75 bps or, for example, 64 users each sending simultaneously 9600 bps. However, this bus occupies only 1 MHz of the cable's bandwidth. Obviously, the capacity could be multiplied by assigning more bandwidth to a single channel so as to create a higher speed bus, or by assigning more channels to multiple busses. In the latter case, there is then an option either to arrange intercommunication between the busses or to isolate some busses from others for privacy or security. The bus is a transparent medium, and data security can be obtained by employing encryption devices between terminal devices and their associated bus interface units.

APPLICATIONS

At the beginning of this paper it was stated that information systems are now being viewed in a comprehensive fashion which includes audio, video, and digital means. Some of the techniques in which these means may be accommodated on a television coaxial cable have been discussed, and it was pointed out that the cable is merely a radio-frequency transmission duct; therefore any technique that can be used with radio can be used on the cable (Figure 4). However, the cable is essentially free of interference and can sustain a high signal-to-noise ratio. Moreover, the cable leads the signal where it is wanted and puts the whole capacity of the 300 MHz bandwidth at the disposal of each one of the using group. The bus architecture and a MITRIX-like time division multiple access organization provides an extremely versatile capability that will support complex multimode systems.

Cable may not be economical for systems where users are tightly grouped, for example, immediately around a computer room or in a number of widely separated tight groups; in these cases point-to-point connections, or a high-speed parallel bus may be appropriate.

For operation over long distances, the cost of laying the cable and installing the amplifiers may exceed that of microwave and other long distance systems. Even if cable is economically feasible, there may be additional problems of obtaining access to existing poles, or obtaining rights of way.

At present, the most appropriate uses of cable appear to lie in large buildings or spatially compact areas where there is a clear need for a wide variety of communications services. Examples are military bases and government office buildings, and in the private sector, hospitals, large insurance company headquarters, highly automated manufacturing complexes and similar organizations. In particular, if the community is already provided with a coaxial cable for entertainment television or for

a surveillance system, the use of the existing cable to support other services becomes even more attractive.

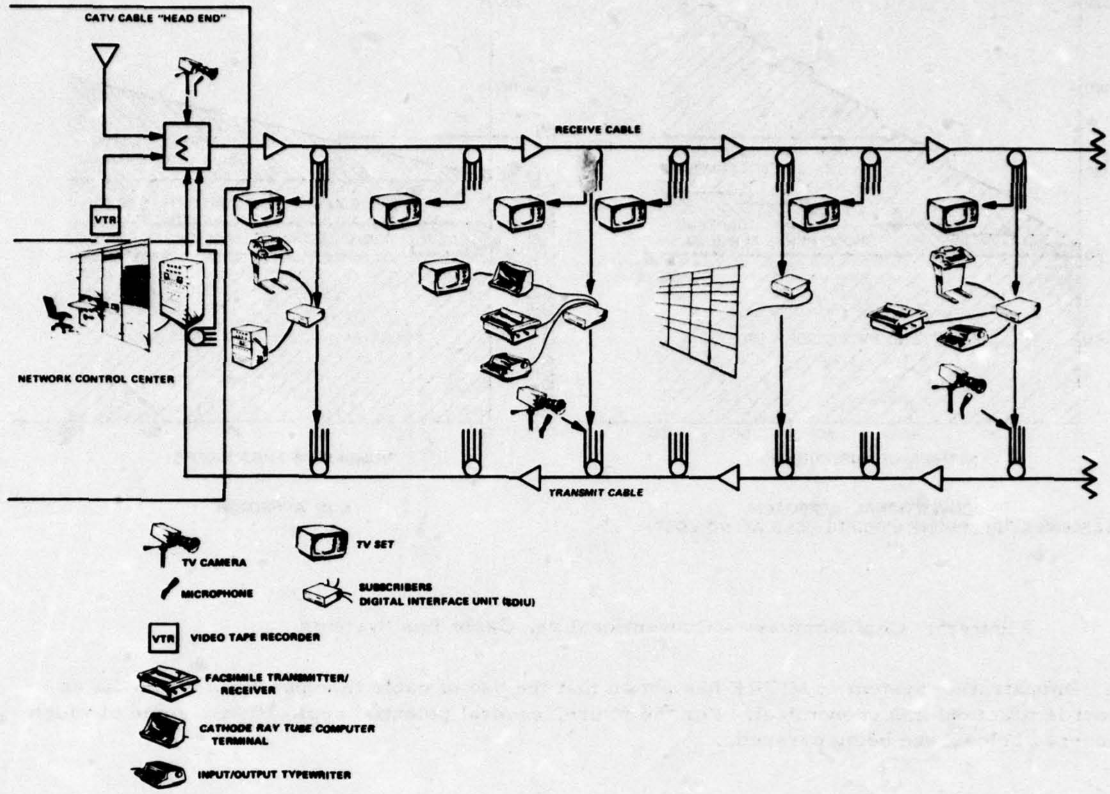


Figure 4. Demonstration System Circuit Configuration

For systems with a very small number of users, the cable bus offers no special advantages. For as few as 20 - 40 users (for example, large computers, distributed terminals or minicomputers, etc.) the economics favor the cable bus. The crossover point will depend upon many factors, including the nature of the existing facilities, the complexity of the interconnections required, and the physical distances involved.

Considering only a low or medium speed digital communications subsystem, and restricting the analysis to communications equipment alone, the expected costs are shown in Figure 5. This diagram shows the effect of adding equipment to a basic front end processor as the number of subscribers is increased. In the conventional approach the additions are line controllers, which handle 16 line increments, modems and couplers. In the cable bus approach a cable distribution system, a network control computer, and bus interface units would be required. For costing, a 1.5 mile cable run has been assumed to be necessary; the modem costs assume a representative mix of data rates on voice-grade telephone lines, but the telephone equipment costs were not included. Also, the cost of providing high data rate communications in the conventional system have been excluded; this capability is, of course, an intrinsic feature of the MITRIX time division design and adds no cost to the cable system. Each installation is, of course, different, but the figure is representative of cost studies performed by MITRE for specific installations and illustrates the attractiveness of the cable bus approach.

Not the least important feature of the bus architecture is that the number of connections and the mass of wire that are employed is much less than those in conventional systems. To connect N terminals to a switch requires 2N connections - to connect them to a bus requires only N+1 (one for the network controller). However, the bus provides unlimited connectivity, which, in the conventional approach requires N(N-1) connections per circuit. In applications such as mobile tactical systems, in which communications must be disconnected before each move, and reconnected afterwards, the reduction in work and in potential error or mechanical damage is obviously advantageous. In the case of the AN/TSQ-91 Control and Reporting Center used by Tactical Air Command, it has been determined that, if the bus architecture had been used, instead of point-to-point communications, the number of circuit connections would be 26 instead of 2800, only 13 cables would have been required instead of 95, and the weight of cables would have been reduced by over two tons.

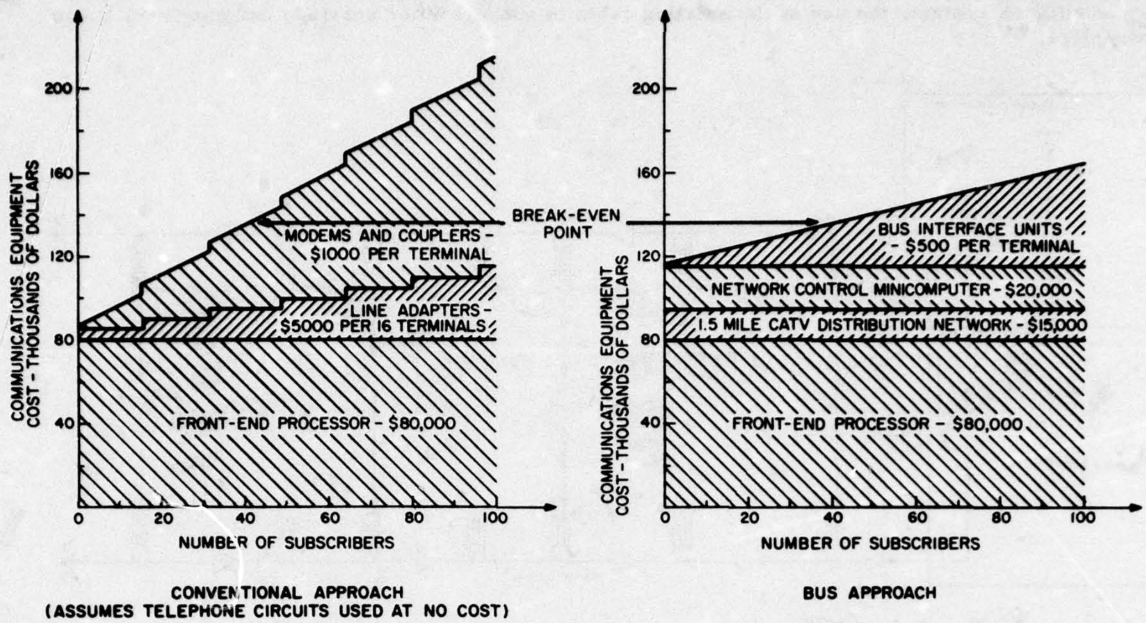


Figure 5. Cost Summary - Conventional vs. Cable Bus Systems

The demonstration system at MITRE has shown that the use of cable to support a multimedia environment is practical and economical. For the future, several potential applications, some of which are discussed below, are being pursued.

Army Base Information Transfer System (ARBITS)

The Army Base Information Transfer System is a major study, supported by the U. S. Army and being conducted by MITRE, which is aimed at the implementation of integrated communications-electronics systems to satisfy the comprehensive information processing and transfer requirements of major Army installations.

It is proposed to continue practical studies of the applicability of the multimedia bus architecture in a test bed facility at Fort Bliss, Texas. Extensive studies of this base indicate that the technology should prove beneficial in several applications areas, such as:

Remote entry and retrieval of digital data.
Forms processing, text editing, correspondence and message distribution.
Data base management, computer programming, and report generation.
Computer-assisted instruction, computer graphics.
Digital telemetry and control.
Computer output microfilm, remote retrieval and update of micro-form data, micropublishing.
Facsimile transmission.
Fixed-frame and broadcast video, interactive access to video, surveillance and teleconferencing.
Secure processing of voice and digital messages.
Switchless telephony.
Linking of telephone system to non-tactical radio nets.

Government Classified Facility

The application of a multimedia cable-based information distribution system has been studied for a classified headquarters facility. The study shows that the expressed requirement for five categories of information transfer can all be supported simultaneously by a single coaxial cable-based network. These comprise:

Digital communications among computer terminals and computer complexes (six IBM 370-168 or larger computers and over 700 terminals).
Document distribution communications - distribution of document images originating in a microfiche scanning facility to 800 remote terminals.
Other digital requirements, including inter-building data transfer and high-speed facsimile.
Communications among 7000 secure telephone sets.
Miscellaneous requirements, such as high bandwidth image transfer, standard television video transmission from the air or tapes.

The digital requirements can be accommodated within less than 10 MHz bandwidth and will operate at 7.4 Mbps with about 50% reserve capacity. This capability is within the current capabilities of low cost digital devices.

University of Vermont

The PRoblem Oriented Medical Information System (PROMIS) that is being developed at the University of Vermont requires, among other things, high-speed digital data communications to ensure that the computer display terminals are responsive to the needs of the medical staff. MITRE has designed and built the prototype modems for a 300 Kbps cable-based system. This is a polling system and the prototypes will be used for proof-of-concept testing of the techniques. Low cost digital circuitry was used for the modems.

Modular Control Center

A study at MITRE, for the United States Air Force, is concerned with the development of a system concept employing interconnected standard modules to support command and control functions. When successful, this approach could reduce the difficulties of acquisition, logistic support, and maintenance of equipment, while providing redundancy which will facilitate the creation of fail-soft systems. Studies have shown that a cable bus acting as the interconnecting vehicle would reduce weight, save set-up time, improve reliability, increase flexibility and facilitate reconfiguration.

CONCLUSIONS

Traditional communications architectures are not adequate to support the transparent, efficient, expandable and easily integrable networks that are needed to exploit new information system concepts. New ideas are needed which will create new structures for distributing information, ideas that will let us tie things together in easier, cheaper and more flexible ways than the usual point-to-point, switched lines.

Computers are no longer limited to expensive central facilities. The minicomputer and the microprocessor are making it possible to put substantial computing power into many more hands, and to bring the computer to the user, rather than the reverse. There is a strong thrust towards distributed processing, dispersing work in specialized machines at many locations. Efficiency and accuracy can be maintained only if these processors can be interconnected economically.

In military systems, the availability of extremely capable but inexpensive digital devices gives the power to develop systems of unprecendented capability and survivability. It will be possible to construct systems from smaller, cheaper building blocks that can be put together in more flexible, more reliable and more survivable systems - but only if we have more flexible communications to support them.

Digital systems are not the whole story. Information systems are being built which exploit the whole range of information handling techniques - not only digital, but video, audio and photographic recording, retrieval and distribution. Such systems require a comprehensive communications system to support them.

Coaxial cable systems, using multimedia bus technology, can provide the required communications support. The bandwidth is sufficient to support extremely demanding information transfer requirements. The technology is well proven; reliable, standard commercial equipment has been available for several years to support high signal-to-noise ratio, interference free transmission. Digital applications are also well proven, but their use has, until recently, been inhibited by cost considerations. However, the recent growth of low cost digital integrated devices has removed the cost barrier - the economics now favor the digital cable for many installations. The costs of digital logic are expected to continue to fall, while there is every indication that installation and maintenance costs, which both have heavy manpower components, will rise. The cost tradeoff should therefore be even more favorable in the future.

TERMINAL ACCESS TECHNOLOGY OF THE 1990s

by CRAIG FIELDS*

INTRODUCTION

I am going to make some predictions about characteristics of terminal access technology of the 1990s. The usual way to predict future technology is to forecast what will be technically possible. This is an unsatisfactory approach, because the number of technological possibilities for the 1990s is truly vast, and most will never reach the market place, much less become popular products in common use. The technology of terminals in the 1990s will be determined by the R&D investment strategy pursued by the computer industry during the next two decades. This investment strategy, in turn, will depend on what the computer industry thinks customers will want in computer terminals of the 1990s. Projecting what people will want in computer terminals of the 1990s would seem to be an even harder job than projecting technological possibilities for the 1990s. Fortunately, terminals actually available twenty years from now will be defined by R&D currently being pursued. It is only necessary to observe what R&D managers currently think people will want in the 1990s, and that is very likely what they will get. As an R&D manager, I am going to present my own views of future desires for computer terminals, and I have not tried to make an opinion survey among my colleagues.

I have been careful to refer to future desires of customers, and not future needs. People are not very good at estimating the most cost effective computer technology to help them do their jobs. Supposed experts usually disagree, and very few scientific experiments have been done relating characteristics of computer systems to improvements in job performance. In those few cases where experiments have been performed, results have been uniformly counter-intuitive. Examples of results include slow time sharing systems are no worse for problem solving than fast time sharing systems; video mediated communications are no better than voice mediated communications for teleconferencing; large vocabularies for speech communication with computers are no better than small vocabularies for speech communication with computers; some times data is better presented as numbers than as graphs; graphical presentation using symbols is frequently better than graphical presentation using colors; less computer-presented data is frequently better for decision making; and so on. The inescapable conclusion is that experimentation is a better approach to designing systems than meditation. Fortunately or unfortunately, it is a lot easier to find out what people want than what they really need, and what they want defines the market place.

Terminal access technology of the 1990s will differ for various jobs and workers. For this paper I will be exclusively concerned with office work, a cerebral function where brains are more important than brawn, thinking more important than sharp eyes and sensitive ears.

I want to describe several kinds of activities that comprise office work, and discuss the ways that computer technology in the 1990s will aid those activities. The most important activity in office work is thinking. It is essential that computer programs be developed to aid people in thinking or, preferably, think for people. The reason for my concern is subtle but important. I believe it inevitable that the problems facing society in the 1990s will require solutions that are beyond the intellectual capability of any individual--person or computer. Both people and computers have fundamental physical limits that they face. For computers that limit is determined by the limited speed of light (only one foot per nanosecond) and atomic limits on the smallest scale computing devices. I cannot easily characterize the forces determining the fundamental limits on human intellect, but I am sure they exist. In any case, if important problems requiring solutions are beyond the capabilities of individuals, they must be solved by groups. We know a reasonable amount about how to combine computers together in order to get increasing computational power. Although all algorithms do not benefit from multi-processor execution, an increasing number can be speeded up. In contrast, we know almost nothing about how to combine people together in order to get increased intellectual capability. A ten-member committee is not ten times as intelligent as any of its members. Since we have not discovered how to combine the intellectual capabilities of people after thousands of years, I have little confidence that we will solve this problem during the next two decades. Computer relacement for human thinking is being vigorously pursued with limited success in the field of artificial intelligence. Computer aids for human thinking are beginning to appear, and include modeling tools (especially heuristic modeling tools), decision making tools, and programs to assist people in plausible reasoning. In summary, increasing scale of problems requires technology for upward scaling of computer power (as a replacement for human thinking) or a new approach to upward scaling or human intelligence (one approach is through computerized teleconferencing systems that actively aid group processes as described below).

Conferencing is going to be an increasingly important function in office work during the next two decades. The expertise of large numbers of different kinds of specialists will be required to solve increasingly complicated social problems. Further, these experts will inevitably be geographically distributed, requiring teleconferencing systems. Computers are not necessary in teleconferencing systems with adequate bandwidth (voice, video, and so on) to mimic face-to-face conferences although computers can reduce the cost of high bandwidth teleconferences by compressing communications using redundancy inherent in human discourse. Computers are necessary, however, as

* DARPA, Arlington, VA., USA

part of teleconferencing systems that contain adequate procedures for controlling the conference (chairman, agenda, recognition of rank of speakers,...) and good recording tools (treating the conference, and parts of the conference, as data), as well as in providing speaker aids such as electronic blackboards and pointers, and electronic slide projectors. Such teleconferencing systems will be aimed at providing high fidelity conferences that appear that almost as if they were taking place in one room. I contend that if high fidelity teleconferencing is a goal of computer technology, we will never achieve a solution to the scaling problem I have discussed--how to combine people together to get increasing capabilities. Rather, I advocate the development of low fidelity teleconferencing, where the communications stream is purposely distorted by computers in teleconferencing system in order to improve communication and group problem solving. I want to illustrate this concept with ten short examples.

(1) People write and speak with a distinctive style that is as idiosyncratic as a fingerprint. It is likely that people can better understand communications in their own style than in anyone else's style. Computers in a teleconferencing system could transform communications from one person's style to another, thus aiding the group decision making process.

(2) An argument can be presented in any number of ways, and the rhetoric selected will depend on the point to be made and the audience that is listening. A computer program could serve to adivse speakers in a teleconference on the best form of argument to use in order to achieve concensus.

(3) Jargon and abbreviations are an invaluable communications aid within a group, but raise a significant barrier to communications between groups. An abbreviator/deabbreviator program within a teleconferencing system could speed up communication between individuals sharing concepts in common (and even create abbreviations for frequently used phrases) and could facilitate communications among strangers.

(4) A computerized diplomat could detect insults, snide remarks and profanity in the communications during a conference, and filter them out in order to aid the group problem solving process. This is a function usually performed by translators at international negotiation sessions.

(5) When people deceive, they reveal themselves through various clues. These include physiological indicators, nonverbal behavior, and changes in communications style. Computer programs, working for individual members of a teleconference, could alert them when they receive potentially deceptive communications from others at the meeting.

(6) Some people talk rapidly and others talk slowly. It is reasonable to believe that fast speakers are fast listeners, and slow speakers are slow listeners. A teleconferencing system can serve like the transmission in an automobile to speed up and slow down communications among different members of a conference.

(7) Most of the contents of messages and briefings are intuitively obvious, and only a small percentage of the communications are devoted to new, surprising facts, insights or opinions. A computerized counter-intuitive filter would serve an important function in allowing the unpredictable material to reach a member of a teleconference, while relieving him of the burden of listening to material he would otherwise expect anyway.

(8) Psychologists have developed a number of rules for managing groups involved in brainstorming. These rules could include suppressing criticism, thinking through metaphors, generating as many ideas as possible, and so on. Such rules could be imposed by a teleconferencing system to aid the creative process.

(9) Each member of a decision making group has his own decision model. That model could be revealed during a teleconference to aid concensus seeking by identifying areas of significant differences of opinion and quickly eliminating areas of agreement.

(10) Beautiful and handsome conference participants who are well groomed are more credible and believable than their less fortunate colleagues. A video based teleconferencing system could include picture processing software that would beautify conference participants, thus aiding them in stating their case.

Document production is an exceedingly mundane but central function of office work. Computerized editors, spelling correctors, grammarizers and format neateners are inevitable developments in computerized office work of the 1990s. Documents themselves will change, becoming multi-media vehicles mixing text, voice, printing, script, line drawings, and half-tone pictures.

Electronic mail is a nonreal-time analog of teleconferencing, and is now available in limited use. Nickname addressing, forwarding, forced routing, and importance alerting and prioritizing are here now. Functions such as redundancy filtering to eliminate unnecessary repetition of messages will be available during the coming decades.

Briefings are a special form of conference that play an essential role in office work. The briefer may be viewed as a guide, leading the audience through a data territory with the goal of convincing the audience to adopt a point of view. A good guide modifies his route, adapting to the responses of the audience. In the 1990s this will

be facilitated because audiences will be instrumented to appraise speakers of their reactions and briefings may be automatically personalized to specific individuals in the audience.

Library searching is achieving increasing importancy in office work as larger and larger computerized data bases become commonly available. In the 1990s, the structure and directory of data bases will be personalized for individuals and made congruent with human memory. More intelligent data base systems may be expected that accept complex, short high level questions from the office worker and return brief, high level answers.

The office workers that I am interested in are decision makers and their staffs. For these individuals, the value of their decisions, mistakes, and opportunities seized or lost is far greater than the cost of the computer systems they will use in solving problems. The importance of their work coupled with the decreasing cost of electronic devices based on improvements in LSI technology and mass production assure that exceedingly fancy terminals will be cost effective for selected office workers in the 1990s.

My approach to defining desires for terminal access technology in the 1990s is based on the observation that computerization of office function has caused big improvements in a number of areas, but has also lead to significant losses in capability because of some characteristics of computer terminals. People are now overcoming their initial infatuation with automation. I believe that the trend of the coming decades will be to regain loss capabilities in an electronic context--if you like, to reinvent the last four thousand years of office work. I do not maintain an anti-technology position. New advances always result in some losses. Computerization is unique in that we can regain what has been lost.

I have identified three areas in which new computer terminals for the 1990s will help us to regain old office capabilities. We will regain some of the advantages of pre-computer communications, some of the special advantages of working with paper, and some of the advantages of the work environment before computer terminals.

THE REBIRTH OF ENRICHED COMMUNICATIONS

At this time there are very few possibilities for man-computer interaction. The technologies available include typing; drawing and pointing with a stylus; and color alphanumerical or graphic displays. Simple trends of the near future are to increase the size of the display and to replace today's fixed alphabet keyboards with dynamic adaptive keyboards presenting a changing alphabet of possibilities depending on the kinds of things people are typing at any particular time.

Computers are getting to be more like people. The field of artificial intelligence is always ten years behind its aspirations and twenty years behind its publicity, but it is making progress. As computers become smarter and more like humans, people will want to regain the enriched communications possibilities they now enjoy with other humans but not with computers. As computers become smarter, they will be able to make good use of new channels of communication.

What channels of communication do people use. Verbal, overt channels include writing and printing; speaking; drawing and pointing with fingers as well as a stylus; painting; and gesturing and moving, as in pantomime. Non-verbal covert channels of communication include eye motion, eye expression, and pupil size; speech clues like pausing and accents; graphology of printing and script; facial expression, "type of face", make-up, moustaches and beards; blushing; sweating and body odor; small body movements; posture; muscle tension and shivering; and social indicators such as dress and haircut. It is technically possible to communicate all of these types of information to computers right now, but we do not know how to write computer programs to interpret and make use of the information.

In addition to these modes of communication currently in use between people and potentially used, through terminals, between people and computers, I cannot resist describing a new alternative. Human brain waves, the electroencephalogram or EEG, is easily measured electronically. It is not a communication channel used by people, but it could be used between people and computers. It would have the greatest communication bandwidth of any communication channel, the shortest latency, the most comprehensive code (including messages otherwise unavailable), the easiest transmission (in energy terms), and would require the least training. Unfortuantely, the EEG has a terrible signal-to-noise ratio. Nevertheless, I have been supporting some research to try to learn the EEG code and use the EEG as a communications channel between men and computers. We have made surprising progress and are now able to get useful information from the EEG about people's attention (how much, to what), fatigue, processing load and reserve, thoughts of specific words, connotation in communication, and surprise, to give a few examples. The technology for doing this is easy, and the algorithms for interpreting the EEG have been drawn from everyday statitistics. EEG can be measured from electrodes imbedded in the headband of hi-fi speakers and magnetic brain waves can be measured without any electrodes by placing an appropriate sensor near the head. This raises the specter of covert mind reading, but I assure you that is not a real danger. Magnetic particles in the hair and clothing produce much larger fields than the brain, and in order to pick up magnetic brain waves it is necessary to demagnetize the head and become magnetically naked.

In summary, I believe terminal access technology of the 1990s will make use of a rich assortment of verbal and nonberbal channels of communication so that man-computer communication will be as rich as man-man communication is now.

REGAINING THE ADVANTAGES OF PAPER

Paper is a marvelous medium for office, work, and we have lost a number of its advantages with the advent of CRT terminals or even with the use of printing terminals. Before computer terminals were used for office work, we purposely or accidentally used many capabilities of paper. Fortunately, these can be regained through properly designed terminals of the 1990s.

It is easy to include important types of information on paper, such as annotations and logos. These are usually lost in computerized documents.

In addition to overt means of representing information, paper documents contain a number of important covert clues. These include the shape of text; the color of text and paper; the type of style of annotation, including writing e.g. printing or script containing their own clues, rubber stamps, stickers and seals; water marks; additions to text showing temporal information; paper documents as an original or a copy; the type of binding, such as paper clips, staples, glue, or sewing; perfumed paper; the weight and texture of paper; or the existence of covers or end papers. This list does not exhaust the clues in paper documents, but it gives some indication of their wealth and variety.

Paper can be manipulated in a number of ways. It can be placed in a particular location, stacked, bound, cut and pasted. Stacks can be flipped or looked at end-on.

These manipulations that are possible with paper allow people to retrieve paper documents using geography as well as symbols. I locate documents by where they are in my office--in specific piles on particular places in my bookshelves, or in particular locations in my files. The ability to manipulate paper allows me to get close to the document of choice, and the use of paper clues allow me to zero in on the particular document I am trying to find. This geometric or geographic approach to paper document management utilizes human place memory. Human place memory is very good, indeed. One standard technique to teach people to have better memory is to suggest that they imagine a room, and place things that they want to remember at different places in the room. From a Darwinian evolutionary viewpoint it is easy to see why human place memory might be considerably better than human symbolic memory.

Fortunately, computer technology makes it possible for us to have the best of both worlds. It is easy to imagine terminals that would allow us to enjoy the advantages of paper as well as the advantages of electronic storage. I want to describe a mythical but technically feasible data management system based on human place memory. It is a data management system for storing text and facts geographically as well as symbolically. I will say nothing more about the symbolic aspects of the storage system, since they are thoroughly conventional. With the geometric version of the system, information is stored at particular places in a two-dimensional or three-dimensional space simulated by the computer storing the information. The screen of the user's terminal is considered a window on that space and the window can move through space, controlled by movements of the hands, eyes, and body of the user. Movements include translation, zooming, and rotation. Navigation can be accomplished in a variety of ways based on a grid gradient throughout the space, a color gradient throughout the space, colored regions, and various landmarks and roads. Information at various places can be stacked just like documents, and can be given all of the previously described paper clues that are basically visual. With such as system it would be possible for the user to enter data, retrieve data, move data, flip through data, search through a pile of data, and so on. Data in the space could be reorganized to take on a personalized view, and yet maintain symbolic links to facilitate sharing and communication among users. This system is within the current state-of-the-art and is under construction. Terminals of the 1990s will have to support the facilities demanded of such geographic data management systems in order to regain a number of the advantages of paper.

Regaining the advantages of the paper will increase communication bandwidth on particular subjects: ease document retrieval; ease document construction; improve authentication of the source of text; and provide new communications about the sender, the sender's view of the subject; and the sender's view of the recipient.

REVIEWING THE WORK ENVIRONMENT

With current computer terminals, office work is done at the computer terminal. The restricted computer terminal work environment is very diferent than pre-computerization work environments, and distinctly worse. The older non-electronic alternative is to have specific work places considerably larger than a terminal; and to be able to work anywhere. Work places would include a desk, filing cabinet, tables, blackboard, bulletin board, conference table, and so on. The ability to work anywhere stems from the portability of pencils, paper, books, and so on.

The concept of specific work places may appear to be in conflict with the idea of working anywhere. That is the case, and yet both possibilities have advantages. The work place carries with it the connotation of the thinking cap--an environment for accomplishment; and yet has a richness not shared by conventional computer terminals.

The possibility of portable work is also highly desirable, and makes it possible to grasp targets of opportunity, such as a pleasant day in the park. Terminal access technology of the 1990s will have two incarnations: first, as an electronic place for doing work, far exceeding current terminals in richness of interaction; second, as portable terminals that are always available to the user for work in any location. The electronic place may be thought of as a terminal that is a room--a multichannel, multi-model terminal that knows the location and attention of the user, and is an active partner with the user in solving problems. The major functional, rather than emotional, advantage of the electronic place is in the explicit use of redundancy. Current computer systems, including computer terminals, avoid redundancy like the plague. However, it is redundancy in the environment that makes it possible for imperfect human beings to function. An environment without redundancy, such as those constructed with modern computers, place a heavy burden on people and inevitably lead to errors.

The construction of a portable computer terminal will not be a significant challenge by the 1990s, due to new batteries, low power flat displays, ultra-low power LSI components, and a variety of I/O means as previously described. Communications may be provided by packet radio or by other means. Although the conventional idea of a portable terminal is the most likely to be desired and hence to occur, more exotic possibilities are conceivable. Implantable terminals will be technically feasible, but probably feared. However, almost as good, terminals may be imbedded in clothing. Electronic clothing could contain a distribution of components connected by an optical packet-switched bus and might be powered thermoelectrically, using threads of dissimilar metals sewn into the fabric. With such a suit, a computer user of the 1990s would have his terminal with him everywhere, or at least everywhere he would want it.

SUMMARY

Regaining the advantages of human communication, paper as an office medium, and intensive and accessible work environments will be a high priority for future purchasers of computer service, and is currently a high priority for the R&D of computer terminals of the 1990s.

ACKNOWLEDGEMENTS

A number of the ideas in this paper were generated during spirited discussions with;

R. Anderson
J. Licklider
J. Markowitz
N. Negroponte

IMPLICATIONS OF FUTURE DEVELOPMENTS IN COMPUTING TECHNOLOGY

Frank V. Wagner
Senior Vice President
Informatics Inc.
21031 Ventura Boulevard
Woodland Hills, California 91364

SUMMARY

A review is made of cost/performance trends in computing technology, including hardware elements and software development. The resulting impact on computer systems architecture is briefly covered. Particular emphasis is placed on the trend toward decentralization. The evolution of the development of distributed computing and the use of modern dedicated small computers is examined and analyzed.

An explanation is given of Grosch's Law, which has described the relationship of computing power and computer price. The possibility is explored that recent developments may invalidate the conclusion based on Grosch's Law (hitherto commonly accepted) that any enterprise should get its data processing done on the largest possible computer that the enterprise can afford to acquire, or to which it can afford to buy access. Conclusions are drawn that a radical revolution in the use of computers has begun, and that in the future, a new "Principle of Decentralized Computing" may replace Grosch's Law.

INTRODUCTION

A while ago, as I was skimming through the Wall Street Journal, I was approached by a philosopher friend of mine. I commented on the interesting item I had been reading about the remarkable growth of the consumer power tool industry. It seems that astonishing records are being set for the sale to home-owners of one-quarter inch electric drills. I expressed my surprise that there were so many people who wanted one-quarter inch drills. My friend observed, "I don't believe there is any real demand for them." When I asked how he could come to such a conclusion in the face of the evidence, he pointed out that, "People do not necessarily want one-quarter inch drills. The things they really want are a great many one-quarter inch holes."

This paper will attempt to look into the future of computing from such a point of view. Computer professionals, in their saner moments, will acknowledge that computers are only tools to enable us all to achieve our real objectives. But unfortunately, computer professionals usually talk and act as though the computer *is an end in itself. However,* the realization is slowly coming into focus that we don't really want one-quarter inch drills; we just want a lot of one-quarter inch holes.

Fortunately, technological improvements which are driving down the cost of computing will make it possible for users to approach the subject a little more rationally than we have in the last twenty years. Therefore, I will first review those technological improvements and give you a glimpse of the cost-performance trends for hardware and for software. Then I will show you the impact that these trends are having, and apparently will have, on the architecture of information processing systems. I will place particular emphasis on the trend toward decentralization. I will trace the evolution of the use of dedicated small computers and of the development of distributed computing. Next, I will explain Grosch's Law, which has described the relationship of computing power and computing price. The possibility will be explored that recent developments may invalidate the conclusion, based on Grosch's Law (and thus, commonly accepted) that any enterprise should get its data processing done on the largest possible computer that the enterprise can afford to acquire, or to which it can afford to buy access. Finally, I will draw the conclusion that a radical revolution in the use of computers has begun, and that in the future, a new "Principle of Decentralized Computing" should replace Grosch's Law.

COST-PERFORMANCE TRENDS IN COMPUTING

GENERAL

Almost any one of the references to this paper will contain the opinion that the costs of all elements of computing are declining, some more rapidly than others, and that, at the same time, performance is increasing. From such generalities, many opinions are formed, and consequences deduced. However, if you ask yourself the question, "How much has cost declined of any particular element of the cost of doing computing?" it becomes a very difficult question to answer, and thus, one begins to doubt the validity of some of the conclusions. I am a great believer in learning from the thoughts of men of demonstrated intellectual capacity. One of my favorite quotations is from the great British physicist of the nineteenth century, William Thompson, Lord Kelvin:

> "When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science."

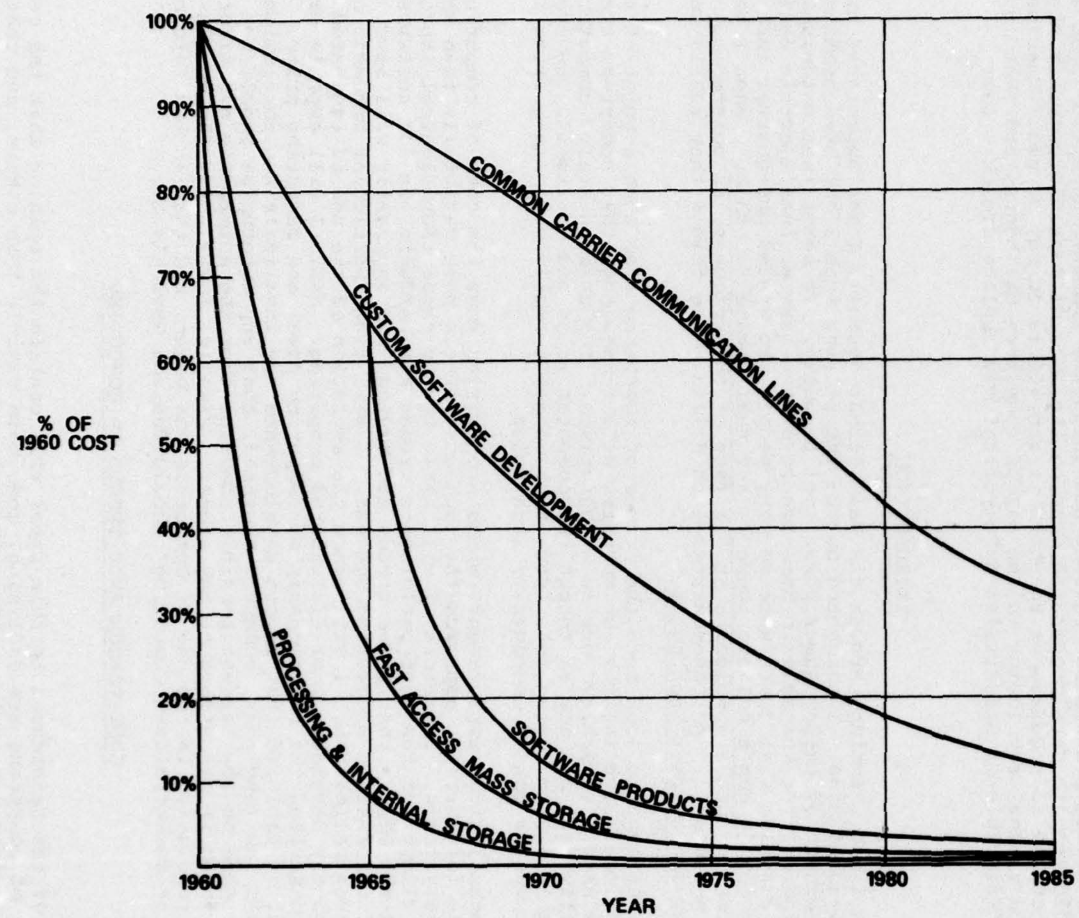


FIG. 1 – RELATIVE COSTS

Consequently, I made an effort to quantify these cost trends.

I immediately ran into several problems of reconciling the available data. No authors approached the matter from the same point of view. None of the performance data was expressed in comparable units. Some authors concentrated primarily on the past, (Ref. 1,2,3,4,5,6,7,8); others, primarily projected into the future (Ref. 9,10,11,12,13). Very few authors defined adequately the makeup of the sub-system they were describing. For example, is a CPU (Central Processing Unit) only the processing logic? Does it include high-speed internal storage? Does it include its power supplies? In many cases, the meaning of "cost" was not defined. Was it the cost of manufacturing, or the ultimate price to the consumer? Since the latter is more meaningful in the context of this paper, I have adopted as my standard.

But that brought me right up against another problem. Withington, in Ref. 10, makes it clear that, after the end of any decade, you simply cannot buy a thing directly comparable to the thing you bought at the beginning of the decade. For example, useful computer systems with only 2K of internal high-speed storage were normal in 1955, but by 1965 they were not generally offered for sale. This situation gets completely unmanageable when one looks at software costs. A very acceptable software program in 1965 would be totally unacceptable in 1975. The 1965 model would have been developed without adequate specifications; it would be prohibitively hard to maintain; its documentation would be rudimentary and inadequate; its testing would not have been subject to modern standards of quality assurance, etc., etc. In spite of such obstacles, I persisted in my efforts at synthesizing the data. The brilliant mathematical logician, Alfred North Whitehead, says: "The art of progress is to preserve order amid change, and to preserve change amid order." In humble emulation of Whitehead and Kelvin, I synthesized, in Figure 1, all the data available to me. In all cases, I show the cost to the user for an element of a computing system relative to its cost in 1960. I believe that, with a few exceptions noted below, I have adequately reconciled most of the anomalies.

HARDWARE

I have selected for study three of the physical elements that go into a computing system: (1) Processing and Internal Storage (to represent the cost of raw computing power), (2) Fast Access Mass Storage (to represent the cost of having all of the data required rapidly available for processing) and (3) Common Carrier Communication Lines (to represent the costs of using a computer from a remote location).

To study "Processing and Internal Storage" I have used the cost of what is commonly called a CPU (Central Processing Unit). It includes all the logic necessary to interpret and execute the instructions to the computer, and the high-speed internal storage in which resides the program and the data currently being processed. It also includes such appurtenances as the power supply and the cabinet. The data was obtained primarily from Ref. 5,10, and Auerbach's paper in Ref. 14. Costs for such CPU's from 1965 on, typically represent the cost of a system with much more powerful logic, and a much greater quantity of internal storage than 1960 CPU's. Hence, the costs shown, particularly for 1975 through 1985, are actually for CPU's much more powerful than the comparable ones of 1960. In spite of this defect in strict comparability, the index of costs shows a spectacular improvement through the years. There is nothing startling about this conclusion — hardly anyone would challenge it. The data are presented for comparison with the other elements of a computing system. 1975 CPU's cost approximately one-half of 1% of those of 1960. The 1985 systems will cost about 20% of the 1975 systems. Clearly, in any analysis of computing costs in the next decade, the cost of the central processor and its internal storage has descended into the noise level.

When one turns to the relative costs of Fast Access Mass Storage, there appears a picture not so widely appreciated. "Fast Access Mass Storage" means the external on-line storage of the computer, in which is held the data (and programs) that could not be held in internal storage, but which requires rapid access for transfer to internal storage for processing. (I do not consider here the more esoteric super-size storage with relatively slow access, currently represented by such devices as the Ampex Terra-bit and IBM 3850.)

Prior to 1960, the most widely used device for this purpose was magnetic tape, inadequately supplemented by small capacity magnetic drums. Beginning in 1960, rotating magnetic disks became more and more prominent, so that eventually, magnetic tapes were largely relegated to the role of off-line and/or archival storage. In the 1980's, it is expected that many of these devices will include semi-conductor technology, CCD (Charge-Coupled Devices, Ref. 11,15) and magnetic domain devices such as bubble memories (Ref. 16). The data were obtained primarily from Ref. 6 and 10. In 1960, the costs were based upon the original 1956 IBM RAMAC, the 1960 IBM 1301, and the 1962 IBM 2311, the first major removable disk pack. The point plotted for 1974 represents the current 200 megabite IBM 3330-II. In order to bring these, and the future more exotic devices, onto a common ground, it was necessary to "normalize" the difference in size of storage and speed of access. For this purpose, I used a "Unit of Capability," defined as the storage capacity in kilo-bites divided by the access time in mille-seconds. Comparable costs, then, are costs per unit of capability. Most of the conventional wisdom that I have come across has agreed that the cost of such fast access mass storage has come down "but not very much." My data show, on Figure 1, that the cost indeed has come down sharply; 1975 costs are only about 2% of 1960 costs, and 1985 costs will be about 10% of 1975 costs.

Since the famous "marriage of computers and communications" (though heralded by Bauer for some years before, the term first appeared in print in Ref. 17), an increasing fraction

of the cost of getting computing done has been the cost of Communications. Cost of "Communications" would include the remote terminal devices, the modems, multiplexors, and concentrators (some of which are on the end of every digital communication line) and the lines themselves. The cost of the lines themselves represent a substantial portion of such costs. The most glamorous portion of line costs are the "long lines" from city to city, now frequently mechanized via satellite. However, what is not commonly recognized is that the local connections (particularly in situations with multiple dispersed terminals where dial-up services are inadequate) frequently represent the largest portion of the cost. Line costs to the user are not necessarily susceptible to advances in technology. Typically, the service must be obtained from a regulated common carrier, whose prices are set by public authorities on the basis of return on investment. High density routes (where unit costs to the carrier are small) frequently command high prices in order to subsidize low density connections to remote hamlets, where the unit costs to the carrier are very large. Thus, the price to the consumer is normally not a function of technology.

My cost data comes primarily from Ref. 3 and 10. On Figure 1 we can see that the reduction of the costs of the communication lines is small, compared with the vastly reduced costs of the other hardware elements of computing. 1975 costs are only 61% of those in 1960; 1985 costs will be about 53% of 1975's. The cost of other elements of the network will, of course, decrease in direct proportion to the amount of electronics that they contain. Nevertheless, it is clear that, of all the elements of computing cost, the least decrease is represented by the communications portion.

SOFTWARE

When one approaches a study of the costs of software, one is keenly aware of Kelvin's dictum, "...when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind.." Professor Brooks has illuminated the difficulty of the problem in his landmark paper, "The Mythical Man Month" (Ref. 18). Nevertheless, a few daring souls have attempted to quantify the problem, principally a team of dedicated researchers at the System Development Corporation. An easily accessible summary of their results is presented by Boehm in Ref. 2. It shows that there is a spread of 1000% between the 10th and 90th percentiles of software productivity, if it is measured by BAL (Basic Assembly Language) instructions per man month. If we take the median of these data, we can tabulate it thus:

Year	1955	1960	1965	1970	1975	1980	1985
BAL Instructions Per Man Month	200	300	450	700	1100	1650	2500

A large part of such improvement in productivity comes from the availability of improved implementation systems — the tools for building software. In 1955 only primitive assemblers were available. Later the first generation of higher order languages appeared; FORTRAN, COBOL and JOVIAL. Today for business application programs, such sophisticated implementation systems as Informatics MARK IV and IBM's RPG II enable programmers to produce a high quality equivalent of BAL instructions from five to fifty times faster than the tools of 1960 permitted. Similar trends are noted in the construction of system programs, where the pioneers, SDC's JOVIAL and Digitec's POPS, have been followed by modern proprietary systems, such as IBM's PL/S, Softek's AED, CAP's CPL/1.

Surprisingly, all this productivity is preserved when converted into cost test per instruction. The cost per man month of a programmer (and his computer test time) capable of producing such an instruction is not growing. It is almost a constant in the range of $3500 - $3800 per man month total cost. Direct labor cost, although growing, has not kept pace with inflation, as shown by Frank in Ref. 3. The increases are compensated for by a big decrease in overhead costs. There has also been a very significant reduction in the costs of machine time used each month by a programmer for testing.

If the above data on instructions per man month are divided by the cost per man month, we obtain a figure for cost per BAL instruction. Relative values thereof are plotted in Figure 1 as the curve labelled "Custom Software Development." The surprising result is that the cost of an equivalent program, comprising 10,000 instructions today is about 28% of its cost in 1960 — and in 1985, it will be less than half of its cost today. But I must repeat, there is no such thing as an "equivalent" program of 10,000 BAL instructions. Today we demand much more; for example, data security, "fail-soft" characteristics, automatic recovery — things which were rarely heard of in 1960. Consequently, it is clear that the cost of developing software has not come down anything like the amount that hardware cost has decreased — indeed, its cost reduction is almost as disappointing as communication costs.

However, another development has taken place, beginning in about 1965 (Ref. 3,19, and 21). In 1960 the concept of a "Software Product" was the dream of a few pioneers. In Ref. 20, Postley described the concept (and the program which was the prototype of Informatics MARK IV, the most successful software product from any supplier except IBM). The concept was simple. Do the development once, but do it in such a way that the identical software could be used by many different organizations. Of course, this has always been done for basic system programs such as language processors, control programs, and access methods. But in 1960, it was deemed impractical to do so for application programs (which represent the major part of software development costs). Stimulated by the success of MARK IV, such extension to application programs began in the late 1960's, and application software products are commonplace today.

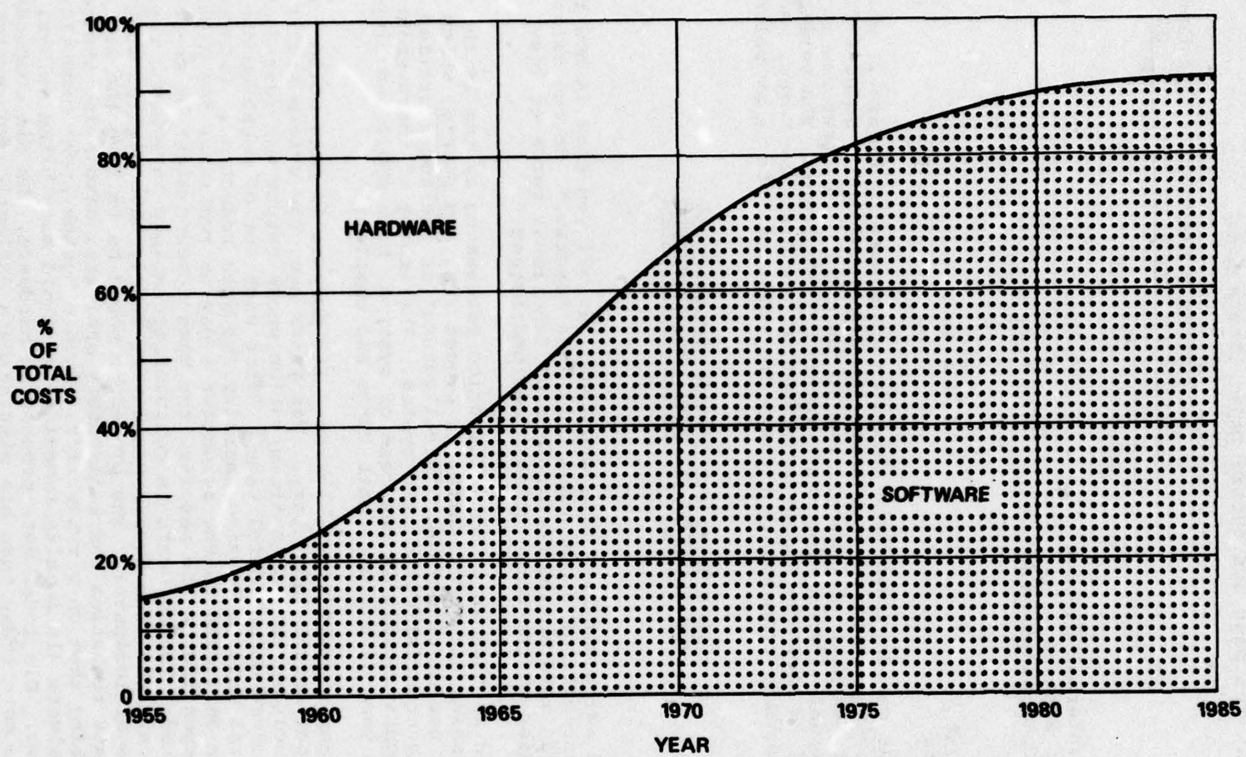


FIG. 2 – DISTRIBUTION OF COSTS BETWEEN HARDWARE/SOFTWARE IN U.S. AIR FORCE

There is no reliable published data on the economics of using software products, and such data would vary widely, depending on the type of product used. I have estimated, however, that the equivalent costs of using a software product had, by 1970, declined to 25% of the cost of a custom developed equivalent system, and by 1975, were certainly no greater than 15%. In some cases, the ratio could be as little as 1%. In Figure 1, I have plotted the most conservative curve, which shows that, using software products, 1975 costs are about 6% of software costs in 1960, and that 1985 costs will be about 33% of 1975's.

SUMMARY OF COST PERFORMANCE TRENDS

Dramatic reductions began in 1960 in cost performance trends for the elements of computing systems. Major cost reductions in computers themselves have been accomplished by 1975, and there will be continuing reductions through 1985. Communication costs, on the other hand, lag far behind all other types of hardware costs. The cost of software development has been progressing nicely, but dramatic improvements can only be achieved by using software products. The past and future improvements are:

	1975 Costs as a Fraction of 1960 Costs	1985 Costs as a Fraction of 1975 Costs
Hardware		
Processing and Internal Storage	0.005	0.20
Fast Access Mass Storage	0.02	0.10
Common Carrier Communication Lines	0.61	0.53
Software		
Custom Development	0.28	0.47
Software Products	0.06	0.33

One conclusion that can be drawn from this is represented by Figure 2, which shows that hardware costs represent an ever decreasing percentage of total budgets for computing. A curve of this sort has been used many times in the past — I first saw it at Informatics in 1963 in an internal planning document prepared by W. Frank. The version shown here was used in Ref. 2 to show the computing budget of the U.S. Air Force. Boehm does not describe therein whether communication costs are included. It is my belief that they are not.

IMPACT ON THE SYSTEM ARCHITECTURE OF INFORMATION PROCESSING SYSTEMS

GENERAL

"System" is a word which can have many meanings. If all you know is how to make screw-threads, then an assembly of a nut and a bolt is a "system." So one satirical definition is: "A system is an interacting combination of functional parts at least one level more complex than the user of the word 'system' can understand."

I am using the word in that way: An information processing system is the entire assemblage of data processing hardware, software, procedures, and people, which are put together to accomplish some useful objective. "Architecture" is the configuration of the elements (with specified characteristics) of the system. The impact of improvements in technology on such systems occurs rapidly in the case of systems both small and simple, and takes a long time in the case of systems both large and complex.

For example, consider a small simple system: Its purpose is data collection and data reduction in an experimental laboratory. Ten years ago, the system architecture would very likely have included analog instrumentation whose output was passed through expensive converters and recorded on a digital tape. After each set of measurements, the tape reel was sent to a central data processing facility for data reduction. Printed results of the experiment were delivered to the scientist a day or two later for plotting, study, and evaluation. Depending on the results, the measurements might be re-recorded or the experiment reoriented. Improvements in computer technology have had a rapid impact on such a system. Today, the transducers would very likely have digital output read directly into an inexpensive minicomputer. The procedure would be to have the minicomputer, on the spot, do the data reduction and rudimentary analysis, displaying the results on a CRT terminal, and printing them on a typewriter. Such a "quick look" capability enables the experimenter to evaluate his results immediately, avoid many blind alleys, and thus, be much more productive. His total data processing hardware, the minicomputer and its terminals, costs no more today than his simpler data collection sub-systems did ten years ago.

In the case of a much larger, more complex system, the impact of technological improvements on system architecture takes a much longer time. For example, consider the business data processing operations of a very large company. Ten years ago, input data was keypunched, and then processed at a central site by the largest available commercial computers, primarily in batch mode. Printed reports were distributed days later. Today, the basic system is not greatly different. Although on-line, transaction-oriented procedures are beginning to be used, the vast majority of today's business processing is done much as it was ten years ago, except that remote terminals may read the cards and print the reports.

TREND TOWARD DECENTRALIZATION

Nevertheless, new trends are beginning to develop. As might be expected, those sub-system components which have had the greatest increase in cost effectiveness have had the greatest influence on architecture. The dramatic reduction of cost in processing logic has created a trend toward incorporating "intelligence," (i.e., processing capability) into the remote terminals. There is a small but definite growth in such intelligent terminals, which are really minicomputers. More and more, the processing which can be done remotely is done remotely, and only that which requires a central computer is transmitted to and from the central site. This concept, known as "distributed processing" is rapidly gaining in popularity. (Ref. 22,23,24,25,26,27, and, especially, 28).

FREE-STANDING MINICOMPUTERS

Ten years ago, if a small sized organization needed data processing, it could not afford to have control of its own data processing system. If it were an independent company, its only choice was to have the work done by a vendor of data processing services. Typically, this was a batch service bureau, which processed the work for a large number of small companies.

If the small unit was part of a large corporation, it had no choice but to send its work to the organization's central data processing facility (Ref. 8). The availability of small minicomputers for business data processing has made possible an entirely different system for getting such work done. (Ref. 29). It is most evident in the small company. In the last five years, due to the introduction of the IBM System/3, and then of the IBM System 32, over 30,000 companies (Ref. 30) have obtained their own computer and do their work under their own control.

In the case of small units within large organizations, such a trend has not yet developed as strongly. Although the same logical forces are operating, they are usually opposed by the vested interest of the management of centralized data processing. (White, in Ref. 31 describes a notable exception.) Distributed processing, described above, is a hybrid system, partially decentralized. It is supported by the management of centralized data processing as long as they can retain control over it. However, since the cost of communication is not coming down as rapidly as all other hardware, there is a strong possibility that it may soon become obvious that the remote processing done under a distributed processing architecture will require a telephone link to a central site only in rare instances. I expect to see such umbilical cords cut with increasing frequency in the next decade.

THE SOFTWARE PROBLEM

At present, the cost of software seems to have mixed effects on the trend toward decentralization. Bookwalter has analyzed this problem, as reported in Ref. 32. On the one hand, there are several factors which seem to indicate that software costs on small decentralized computers should be greater than such costs on large central computers, and thus inhibit the trend. Sophisticated tools for software development on small computers are available in much less variety than on large machines. Good operating systems, with conveniences for the programmer, are rare. With the exception of RPG II on Systems/3 and 32, second generation implementation systems, such as COBOL compilers and only one Data Base Management System, a small version of TOTAL, are just beginning to become available on a relatively few small machines. Third generation implementation systems, necessary for high productivity, such as Informatics MARK IV, are available on only one small computer — the Microdata "REALITY" System, mentioned in Ref. 9.

On the other hand, there are opposing factors which tend to reduce the cost of software development on decentralized minicomputers. Perhaps the most powerful factor is that the computer is specialized. It is not all things to all men — when best used, it tends to be dedicated to a single application or a group of related applications. Under such circumstances, there is no need for a sophisticated operating system with all its mind-boggling complexity. (Patrick's excellent analysis, Ref. 33, discusses how system complexity can increase costs by a factor of four.) Therefore, the programming is straightforward and uncomplicated. Another factor of increasing importance is the accelerating trend toward the availability of software products on minicomputers. As discussed above, software costs can be reduced most dramatically by the use of general purpose software products. The System 32 is better equipped with the application products than any large machine. As these become more available for other minis, the prior inhibiting effect of software development on the trend will disappear.

HARDWARE

The dramatic decrease in the cost of processing logic has made possible a development which has not yet emerged as a trend. But it seems so logically inevitable that I expect to see it emerge within the next five years. The concept addresses the software problem by exploiting the low cost of hardware logic to use existing software. I can best illustrate it by describing a case in which the technique produced dramatic results.

Almost ten years ago, a large organization needed a very sophisticated message switching system. They selected a very powerful, highly sophisticated computer — the best one for the purpose available at the time. The software necessary to implement the system was a major development project. It took a number of years and cost several million dollars. Subsequently, several other organizations acquired the same hardware, and built

on the base of the same software, enhancing it as the years went by. The replacement costs of that software today would perhaps be $5,000,000; it was by far the best available answer to the requirements. In the early 70's, the hardware began to become obsolescent and its cost-effectiveness declined, so that new users were reluctant to acquire such hardware. New modern minis, as we have seen above, could perform the same functions at far less cost. But their use was impractical; <u>the cost of rewriting the software was prohibitive</u>.

So a concept was developed to use such relatively inexpensive hardware for processing the existing software by the well-known technique of hardware emulation. A relatively inexpensive, reliable minicomputer from a reputable manufacturer was selected. Enhancements to the hardware were designed, which enabled the message-switching software to run on the foreign machine. The cost of hardware development and manufacturing was less than 10% of the cost of rewriting the software. The project was successfully completed, and a substitute was developed, highly cost-effective by 1972 standards. By 1980, such a substitute can be developed for 20% of the cost of the 1972 model. (Ref. 34,35).

There is available today a growing inventory of software products for large, expensive computers. I believe that inexpensive hardware enhancements to modern minicomputers will soon enable them to run this inventory of available software. The most cost effective approach will be to dedicate a machine to a specific purpose (See Ref. 33). Today you have on your wrist a computer dedicated to telling you what time it is. Soon you will buy a computer to do only accounts receivable or document accession, or whatever your job is.

GROSCH'S LAW

In the first two decades of the computer age, from 1950 through 1970, a number of generations of machines were developed. As each major computer was offered for sale, it could perform more work for less money than its predecessors. Dr. H.R.J. Grosch, the recently elected president of the Association for Computing Machinery, was the first one to popularize a formal description of this phenomenon. Humorously at first, it became known as Grosch's Law, formulated as follows: "Throughput capacity of a computer is proportional to the square of its price." A number of more formal documented studies of this phenomenon have confirmed that the "Law" was indeed valid for the large computers of that era when operated in batch mode. (Ref. 1,7). (Cynics, however, pointed out, in Ref. 8, that the Law was not fundamental because of the economics of engineering development, but only because of IBM pricing strategy.)

In the same era, there was another phenomenon occurring. The users of such large machines were making greater and greater use of computers. More and more computer applications were being developed, as described in Ref. 33.

Data processing management, *responsible for selection and procurement of computing equipment*, became well aware of the "economies of scale," (Ref.1,7). Since the demand for capacity seemed to be growing without limit, there evolved a corollary to Grosch's Law that I call the "Dogma of Data Processing." It is: "The most economical way to do computing is to acquire the <u>largest</u> computer that the enterprise can possibly foresee a need for; therefore, <u>all</u> computing in the enterprise <u>must</u> be done in one <u>central</u> computing facility." The application of the Dogma made it clear that an application of "Engine Charlie" Wilson's famous quotation: "What is good for General Motors..." was not bad for General Motors Data Processing manager! His prestige, power authority, and salary were also proportional to the size of the installation that he ruled. Hence, there has evolved a strong "union" of data processing managers, who quote the gospel according to Grosch in defense of centralization, and treat as heretics the radicals who would propose decentralization.

THE REPEAL OF GROSCH'S LAW

As the technological developments described earlier began to occur, the price-throughput relationship of Grosch's Law began to disappear. Half in jest, but with astounding foresight, Adams first headlined "GROSCH'S LAW REPEALED" in 1962. (Ref. 36). On the one hand, the very large machines began to encounter some <u>dis</u>-economies of scale. As more and more work was done on a single mainframe, the operating system, in order to sort it all out, got more and more complicated. Operating system overhead began to climb, so that less and less useful work was actually being done. Patrick, in Ref. 33, points out the possibility of such systems growing large enough to fall of their own weight.

On the other hand, as we have seen, the new small cheap minicomputers simply did not follow the law that "throughput capacity is proportional to the square of the price." Although I am unaware of any computations to demonstrate the fact, I believe that the data since 1965 would support Adam's 1962 conjecture (Ref. 36) that: "Throughput performance is proportional to the <u>square root</u> of the price." If his conjecture is true, the cheapest way to get computing done is to get the smallest computer that can perform a particular application. The $8.95 hand-held calculator may be an illustration of the latter "law."

THE REAL REASON FOR DECENTRALIZATION

WHY DO WE USE COMPUTERS?

Everything that I have stated above is really a smoke screen — it obscures the real forces and real trends which will determine the shape of the future. If you listen to data processing people with a critical ear, you will begin to be appalled at the fact that they seem to believe that the computer is an end in itself. Remember my story about the quarter-inch drills and the quarter-inch holes? The comparison is clear — nobody wants computers — people only want to get their work done. If the computer is a tool which can help them get their work done, they ought to use it, and will use it. If the computer gets in people's way, they may be forced by circumstances to use it, but their productivity will be lower than it could be. We must never forget that the work is done by people, not computers.

OPTIMIZATION — LOCAL VERSUS GLOBAL

In most large organizations, data processing costs tend to be somewhere from 0.5% to 2% of total costs. Salaries and other costs of people are between 40% and 90% of total costs. The efficiency of the total enterprise is the efficiency of the global system of people, doing the necessary work, assisted by computers. Yet, at meetings of computer people, the discussions might lead one to believe that all the work was done by computers. Serious papers are published on optimizing the hardware/software sub-system, leading the unwary to believe that such optimization would solve the whole problem. Such local optimization can be useful, but, as is well known, must always be subordinated to the global optimization of the real problem, the total system.

THE EFFICIENCY OF PEOPLE

Just suppose that typewriters had never been invented. All the secretaries in your office are producing letters and reports by writing them out very neatly in longhand with ball-point pens. Now IBM makes a dramatic announcement — it has invented the typewriter! Next morning, at every large corporation, an IBM typewriter salesman has an appointment with the Vice President for Office Services. Let us see what happens, for example, at the General Electric Corporation. The salesman suggests that GE buy a $600 typewriter for each secretary. This is a revolutionary proposal. Purchasing is called in. Internal Consulting is summoned to a meeting. Operations Research is charged with investigating the concept, and conducting a feasibility study. The results show that the average use of a typewriter by a secretary would be 1.1873 hours per day, and that the productivity of secretaries would increase by 325.26%. It is strongly recommended that the corporation convert to typewriters. It is recommended that enough typewriters be procured so that each is loaded 4.7492 hours per day. Allowing for down time and assuming that most overloads can be handled by over-time, each four secretaries will share one typewriter.

The recommendation is adopted. The typewriters are delivered, and training begins. Soon all secretaries are mechanically proficient; Purchasing cancels all orders for ball-point pens; and the system is cut over to document production by typewriter only. Productivity is very low the first week. It is worse the second week, worse the third, and by the end of the month, the backlog of letters and reports has reached alarming proportions. Most executives are spending a good deal of their time consoling tearful secretaries who complain that they cannot get their work done because they cannot get access to a typewriter. An emergency meeting is held. Only two viable alternatives present themselves — go back to the ball-point pens, or get a typewriter for each secretary! Operations Research does a fast study, and concludes that spending four times as much for the typewriters will be paid for many times over by the increase in productivity. The recommendation is accepted, the secretaries live happily ever after. and General Electric increases its dividends.

The moral of the fairytale is clear. Efficiency of people can vary over a wide spectrum. The productivity of programmers has been studied extensively; no general conclusion was possible because productivity varied by a factor of ten between different people, and even when measuring the same person at different times. All data on the efficiency of people show that it can seriously be impaired by frustration — "Nobody lets me get my work done without interference." Conversely, productivity seems to be at its highest when the worker has full control of the tools necessary to do a job. Many of you have had experiences with central data processing that parallel those that the secretaries had with the shared typewriters. You complain of the service. Data processing management writes an authoritative study showing that the service is excellent and very cost-effective (implying that you are the money-waster). You fire back a memo documenting how bad it is. In the meantime, the work suffers. What would happen if you had control of your own computer? If anything went wrong, you would look around for someone to blame, and see no one but your own unit. You would roll up your sleeves and solve the problem.

Of course, the moral of my fairytale would not be applicable if typewriters cost $600,000 apiece, instead of $600. Until ten years ago, computers were too expensive to consider decentralizing control of them. But I submit that today their prices, including software, are approaching the low level where it is foolish to sub-optimize computing costs at the expense of the costs of the people who really do the work.

THE PRINCIPLE OF DECENTRALIZED COMPUTING

I believe that we have entered an era where guidance in acquiring computing power can be derived from a new Principle of Decentralization:

> "If an organizational group, smaller than 30 people, requires computer assistance, it is better for the total enterprise that those people have exclusive use of their own computer — provided that the computer, big enough to do the job properly, will be loaded to over 10% of its capacity."

The underlined words are vital to the application of the principle.

The group must be smaller than 30 people, a number that should decrease in the future as the costs of the computer decrease. It was chosen large enough so that the cost of a dedicated computer would be less than 5% of the total personnel costs. (Surely, adding a computer should improve the efficiency of an operating unit by 5%!) On the other hand, if the group is larger than 30, it probably will have more than one kind of work to do, and thus, the computer would begin to be used for more than one application. Dorn shows, in Ref. 37 that when this happens, you immediately have a small central computing facility, with all its attendant evils, and you lose many of the advantages of decentralization. Thus, the importance of the words, "exclusive use," which cannot happen unless the group is smaller than 30 people.

The computer must be big enough to do the job properly. There are many problems which simply cannot be handled by today's small computers. Nuclear reactor design, numerical prediction of global weather, maintenance of reservations a year ahead for all of the airline seats in the world, instant retrieval from massive data bases, are BIG problems. Such applications are only feasible today on a computer so expensive that is only practical if it is shared by many groups of users. I venture no predictions as to the speed with which technological improvements will catch up with these applications. Up until recently, I believed that data bases would be the last application to become adaptable to a dedicated small computer. But a careful analysis shows that such is not always the case. For example, data bases can frequently be segmented and distributed to the place where the segment is used. "Instant" updating may be unnecessary; Ref. 28 notes that a large insurance policy file has been distributed among branch offices, each of which services a distinct small group of policy holders.

I have recently become familiar with an application involving a medium-sized bibliographic data base — the catalog cards of the Library of Congress. Informatics has developed a system, called MINI-MARC, to increase the productivity of the cataloging staff of a library. Such a library uses, as its source information, the data issued in MARC format for each book by the Library of Congress. However, it is not appropriate to use the information in its raw form for the particular library's own catalog, which is unique because of the unique needs of the library. The cataloging staff eliminates some of the MARC data, and adds additional data. The Informatics MINI-MARC system consists of an inexpensive minicomputer processor, two floppy-disk drives, a cathode ray tube terminal with keyboard, and a typewriter for hard copy output. A part of the system is a data base of all such Library of Congress MARC data on 400 diskettes. Informatics provides a subscription service to keep the data base up to date as it is issued. The cataloger selects the proper diskette and mounts it on one drive. Using the cathode ray tube and its keyboard, he reads the information, deletes from it, adds to it, puts it into exactly the form that he wants in his catalog, and writes that on a diskette on the other drive. Thus, he has two data bases at his disposal in the computer — the original catalog card and his own library's catalog. The diskettes containing his own library's catalog are periodically processed by a large computer at a central site to produce the printed catalogs his library needs. The entire system can be obtained for under $50,000, including all the software to make it work. On a lease basis, the monthly cost would not be greater than the salary of one cataloger. Two such systems are expected to be installed in a Government library in the near future.

The last part of the principle of decentralization speaks to the utilization of such a dedicated computer. It conjectures that the computer should be loaded to over 10% of its capacity. Logically, of course, the loading is immaterial, as long as the productivity of the group is increased enough to pay for the computer. However, I have added this requirement, without any analytical justification for the selection of the 10% level of use, because of my belief that, if it is to be successful, the assistance of the computer to the group should not be for some trivial application, but should participate in the main-stream activity of the group.

CONCLUSIONS

It is now clear that Grosch's Law has been repealed by technological advances. IT IS NO LONGER TRUE THAT THE MOST COST-EFFECTIVE WAY TO DO DATA PROCESSING IS ON A LARGE CENTRAL COMPUTER. However, the current investment of large amounts of money in central data processing installations and in the organizations built to support them, and especially the vested interests of their management and of IBM's manufacturing capability, will delay for many years the inevitable growth of decentralization.

The new "Principle of Decentralization" is a useful guide to evaluating whether to use central computing or not.

It is likely that there is a big future for standard software application products, running on small dedicated computers enhanced by emulation hardware.

REFERENCES

1. Littrell, R.F., "Economies of Scale in the IBM 360 and 370," Datamation, March, 1974, pp. 83-88.

2. Boehm, B.W., "Software and Its Impact: A Quantitative Assessment," Datamation, May, 1973, pp. 48-59.

3. Frank, W.L., "The Second Half of the Computer Age," Datamation, May, 1976, pp. 91-100.

4. Ware, W.H., "Computers and Society and the Technological Setting," RAND Paper P-5094, October 1973.

5. Armer, P., "Computing Aspects of Technological Change, Automation, and Economic Progress," Technology and the American Economy, the report of the National Commission on Technology, Automation, and Economic Progress. Appendix Volume I, The Outlook for Technological Change and Employment, U.S. Government Printing Office, Washington, D.C. Feb. 1966, pp. I-205 - I-232; also published as RAND Paper P-3478.

6. Reports 70c-491-03, Jan. 1974 and 70c-491-04, Apr. 1976,"Datapro 70, The EDP Buyers Bible," Vol. 1.

7. Solomon, M.B., "Economies of Scale and the IBM System/360," Communications of the ACM, June, 1966, pp. 435-440.

8. Oldehoeft, A.E. and Halstead, M.H., "Maximum Computing Power and Cost Factors in the Centralization Problem," Communications of the ACM, Feb. 1972, pp. 94-96.

9. Martin, R.R. and Frankel, H.D., "Forecast of Computer Memory Technology," Comp Con 1974, IEE Catalog No. 74CH0825-OC, pp. 287-290.

10. Withington, F.G., "Beyond 1984: A Technology Forecast," Datamation, Jan. 1975, pp. 54-73.

11. House, D., "CCD vs. RAM for Bulk Storage Applications," Comp Con 1976, Spring Edition, Session 7, IEE Catalog No. CH 1069-4(CR), pp. 60-61.

12. Conley, S.T., "Tape, Disk — What Next?," Comp Con 1974, Session 2, IEE Catalog No. 74CH0825-OC, pp. 6-7.

13. Yasaki, E.K., "The Emerging Microcomputer," Datamation, December, 1974, pp. 81-86.

14. Moneta, J., Editor, "New Horizons in the Computer Field" U. Galil, D. Chevion, W.L. Frank, C.C. Gotlieb, I.L. Auerbach, Contributors, Information Processing Association of Israel, Vol. 13, April 1974.

15. Bowers, D.M., "The Charge-Coupled Device — Is This the End for Core, Drum and Disk?," Modern Data, August, 1975, pp. 54-56.

16. Kaplan, A.R., "Bubble Memories," Modern Data, September, 1975, p. 49.

17. Buaer, W.F., "Computer/Communications Systems — Patterns and Prospects" from "Computers and Communications, Toward a Computer Utility," Gruenberger, R., Editor, Prentiss-Hall, 1968

18. Brooks, F.P., Jr., "The Mythical Man-Month," Datamation, December, 1974, pp.44-52.

19. Bauer, W.F. and Rosenberg, A.M., "Software: Historical Perspectives and Current Trends," Proceedings of FJCC, AFIPS Press, November, 1972.

20. Postley, J.A. and Buetell, T.D., "Generalized Information Retrieval and Listing System," Datamation, December 1962, pp. 22-25.

21. Bauer, W.F., "Software Products in the Next Ten Years," paper delivered at American Institute of Industrial Engineers — Software Conference, July, 1976, pp. 19-21.

22. Editorial Staff Article, "Distributed Computing: A Growing Concept," Infosystems, August, 1975, pp. 32-34.

23. Carren, D.M., "Multiple Minis for Information Management," Datamation, September, 1975, pp. 54-58.

24. Ashenhurst, R.L. and R.H. Vonderohe, "A Hierarchical Network," Datamation, February, 1975, pp. 41-44.

25. Farber, D.J., "A Ring Network," Datamation, February, 1975, pp. 44-46.

26. Wulf, W. and R. Levin, "A Local Network," Datamation, February, 1975, pp. 47-50.

27. Fraser, A.G., "A Virtual Channel Network," Datamation, February, 1975, pp. 51-54.

28. Canning, R.G., "Distributed Data Systems," EDP Analyzer, June, 1976, Vol. 14, No. 6.

29. Marienthal, L.B., "Small Computers for Small Businesses," Datamation, June, 1975, pp. 62-78.

30. Wiseman, T., "Small Business Base to Grow 20%/Year," Computerworld, July 12, 1976 pp. 33-34.

31. White, R.B., "A Case Against Large-Scale Computers," Infosystems, August, 1975, p.34.

32. Yasaki, E.K., "The Mini: A Growing Alternative," Datamation, May, 1976, pp. 139-142.

33. Patrick, R.L., "Decentralizing Hardware and Dispersing Responsibility," Datamation, May, 1976, pp. 79-84.

34. Ogdin, J.L., "Other Microcumputer Chips — The CPU Is Not Enough," Modern Data, February, 1975, pp. 43-47.

35. Garen, E.R., "Applying Microprocessors and Microcomputers — Technology is Useless Unless It Works," Moden Data, February 1975, pp. 54-57.

36. Adams, C.W., "Grosch's Law Repealed," Datamation, July, 1962, pp. 38-39.

37. Dorn, P.H., "The Trouble With Minis," Datamation, May 1976, p. 82.

38. Turn, R, "Air Force Command and Control Information Processing in the 1980s: Trends in Hardware Technology," RAND Corporation, Report No. R 1011-PR, October, 1973.

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP-207 ✓	ISBN 92-835-0183-7	UNCLASSIFIED

5. Originator	Advisory Group for Aerospace Research and Development North Atlantic Treaty Organization 7 rue Ancelle, 92200 Neuilly sur Seine, France ✓
6. Title	ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION SYSTEMS ✓
7. Presented at	the Technical Information Panel Specialists' Meeting held in Arlington, VA, USA, 20–21 October 1976.

8. Author(s)	9. Date
Various	December 1976
10. Author's Address	**11. Pages**
Various	162

12. Distribution Statement	*This document is distributed in accordance* with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Information retrieval	Data processing
Information systems	Small scale computers
Data retrieval	Information retrieval effectiveness

14. Abstract

The ultimate test of any information system is whether or not the user can quickly, easily and comprehensively obtain relevant information stored therein. As data bases expand and multiply, retrieval technology must advance if the user is to benefit fully from this increased information potential. For this reason, three of the four sessions of this meeting, covering minicomputers, bibliographic data base sharing and future technological advances, explore in some depth areas critically important to the user's need to access an increasing number of scientific and technical information collections with greater ease and efficiency.

The other session was somewhat differently oriented and addresses numerical data bases, a relatively new and rapidly expanding addition to information systems. Presentations include not only the development of such data collections but several applications as well.

AGARD Conference Proceedings No.207
Advisory Group for Aerospace Research and Development, NATO
ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION SYSTEMS
Published December 1976
162 pages

The ultimate test of any information system is whether or not the user can quickly, easily and comprehensively obtain relevant information stored therein. As data bases expand and multiply, retrieval technology must advance if the user is to benefit fully from this increased information potential. For this reason, three of the four sessions of this meeting, covering minicomputers, bibliographic data base sharing and future technological

P.T.O.

AGARD-CP-207

Information retrieval
Information systems
Data retrieval
Data processing
Small scale computers
Information retrieval effectiveness

AGARD Conference Proceedings No.207
Advisory Group for Aerospace Research and Development, NATO
ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION SYSTEMS
Published December 1976
162 pages

The ultimate test of any information system is whether or not the user can quickly, easily and comprehensively obtain relevant information stored therein. As data bases expand and multiply, retrieval technology must advance if the user is to benefit fully from this increased information potential. For this reason, three of the four sessions of this meeting, covering minicomputers, bibliographic data base sharing and future technological

P.T.O.

AGARD-CP-207

Information retrieval
Information systems
Data retrieval
Data processing
Small scale computers
Information retrieval effectiveness

AGARD Conference Proceedings No.207
Advisory Group for Aerospace Research and Development, NATO
ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION SYSTEMS
Published December 1976
162 pages

The ultimate test of any information system is whether or not the user can quickly, easily and comprehensively obtain relevant information stored therein. As data bases expand and multiply, retrieval technology must advance if the user is to benefit fully from this increased information potential. For this reason, three of the four sessions of this meeting, covering minicomputers, bibliographic data base sharing and future technological

P.T.O.

AGARD-CP-207

Information retrieval
Information systems
Data retrieval
Data processing
Small scale computers
Information retrieval effectiveness

AGARD Conference Proceedings No.207
Advisory Group for Aerospace Research and Development, NATO
ADVANCEMENTS IN RETRIEVAL TECHNOLOGY AS RELATED TO INFORMATION SYSTEMS
Published December 1976
162 pages

The ultimate test of any information system is whether or not the user can quickly, easily and comprehensively obtain relevant information stored therein. As data bases expand and multiply, retrieval technology must advance if the user is to benefit fully from this increased information potential. For this reason, three of the four sessions of this meeting, covering minicomputers, bibliographic data base sharing and future technological

P.T.O.

AGARD-CP-207

Information retrieval
Information systems
Data retrieval
Data processing
Small scale computers
Information retrieval effectiveness

advances, explore in some depth areas critically important to the user's need to access an increasing number of scientific and technical information collections with greater ease and efficiency.

The other session was somewhat differently oriented and addresses numerical data bases, a relatively new and rapidly expanding addition to information systems. Presentations include not only the development of such data collections but several applications as well.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Arlington, VA, USA, 20–21 October 1976.

ISBN 92-835-0183-7

advances, explore in some depth areas critically important to the user's need to access an increasing number of scientific and technical information collections with greater ease and efficiency.

The other session was somewhat differently oriented and addresses numerical data bases, a relatively new and rapidly expanding addition to information systems. Presentations include not only the development of such data collections but several applications as well.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Arlington, VA, USA, 20–21 October 1976.

ISBN 92-835-0183-7

advances, explore in some depth areas critically important to the user's need to access an increasing number of scientific and technical information collections with greater ease and efficiency.

The other session was somewhat differently oriented and addresses numerical data bases, a relatively new and rapidly expanding addition to information systems. Presentations include not only the development of such data collections but several applications as well.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Arlington, VA, USA, 20–21 October 1976.

ISBN 92-835-0183-7

advances, explore in some depth areas critically important to the user's need to access an increasing number of scientific and technical information collections with greater ease and efficiency.

The other session was somewhat differently oriented and addresses numerical data bases, a relatively new and rapidly expanding addition to information systems. Presentations include not only the development of such data collections but several applications as well.

Copies of papers presented at the Technical Information Panel Specialists' Meeting held in Arlington, VA, USA, 20–21 October 1976.

ISBN 92-835-0183-7

AGARD

NATO OTAN

7 RUE ANCELLE · 92200 NEUILLY-SUR-SEINE

FRANCE

Telephone 745.08.10 · Telex 610176

DISTRIBUTION OF UNCLASSIFIED AGARD PUBLICATIONS

AGARD does NOT hold stocks of AGARD publications at the above address for general distribution. Initial distribution of AGARD publications is made to AGARD Member Nations through the following National Distribution Centres. Further copies are sometimes available from these Centres, but if not may be purchased in Microfiche or Photocopy form from the Purchase Agencies listed below.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordonnateur AGARD – VSL
Etat-Major de la Force Aérienne
Caserne Prince Baudouin
Place Dailly, 1030 Bruxelles

CANADA
Defence Scientific Information Service
Department of National Defence
Ottawa, Ontario K1A OZ2

DENMARK
Danish Defence Research Board
Østerbrogades Kaserne
Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92 Châtillon sous Bagneux

GERMANY
Zentralstelle für Luft- und Raumfahrt-
dokumentation und -information
Postfach 860880
D-8 München 86

GREECE
Hellenic Armed Forces Command
D Branch, Athens

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio del Delegato Nazionale all'AGARD
3, Piazzale Adenauer
Roma/EUR

LUXEMBOURG
See Belgium

NETHERLANDS
Netherlands Delegation to AGARD
National Aerospace Laboratory, NLR
P.O. Box 126
Delft

NORWAY
Norwegian Defence Research Establishment
Main Library
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Direccao do Servico de Material
da Forca Aerea
Rua de Escola Politecnica 42
Lisboa
Attn: AGARD National Delegate

TURKEY
Department of Research and Development (ARGE)
Ministry of National Defence, Ankara

UNITED KINGDOM
Defence Research Information Centre
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RE

UNITED STATES
National Aeronautics and Space Administration (NASA),
Langley Field, Virginia 23365
Attn: Report Distribution and Storage Unit

THE UNITED STATES NATIONAL DISTRIBUTION CENTRE (NASA) DOES NOT HOLD STOCKS OF AGARD PUBLICATIONS, AND APPLICATIONS FOR COPIES SHOULD BE MADE DIRECT TO THE NATIONAL TECHNICAL INFORMATION SERVICE (NTIS) AT THE ADDRESS BELOW.

PURCHASE AGENCIES

Microfiche or Photocopy
National Technical
Information Service (NTIS)
5285 Port Royal Road
Springfield
Virginia 22151, USA

Microfiche
Space Documentation Service
European Space Agency
10, rue Mario Nikis
75015 Paris, France

Microfiche
Technology Reports
Centre (DTI)
Station Square House
St. Mary Cray
Orpington, Kent BR5 3RF
England

Requests for microfiche or photocopies of AGARD documents should include the AGARD serial number, title, author or editor, and publication date. Requests to NTIS should include the NASA accession report number. Full bibliographical references and abstracts of AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR),
published by NASA Scientific and Technical
Information Facility
Post Office Box 8757
Baltimore/Washington International Airport
Maryland 21240, USA

Government Reports Announcements (GRA),
published by the National Technical
Information Services, Springfield
Virginia 22151, USA

Printed by Technical Editing and Reproduction Ltd
Harford House, 7–9 Charlotte St, London W1P 1HD

ISBN 92-835-0183-7